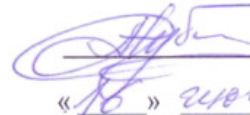


Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт космических и информационных технологий
Кафедра «Информатика»

УТВЕРЖДАЮ
Заведующий кафедрой
«Информатика»

 /Рубан А.И.
«16» июня 2016 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ


Формирование мультилингвистического информационно-
терминологического базиса для адаптивных обучающих систем
27.04.03.02 «Системный анализ данных и технологии принятия решений»
27.04.03 «Системный анализ и управление»

Научный руководитель

 15.06.16, доцент, К.Т.Н.
подпись, дата должность, ученая степень

А.А. Даничев

Выпускник

 15.06.2016
подпись, дата

А.М. Мартоян

Рецензент

 _____
подпись, дата должность, ученая степень

А.Л. Двинский

Красноярск 2016

РЕФЕРАТ

Выпускная квалификационная работа по теме «Формирование мультилингвистического информационно-терминологического базиса для адаптивных обучающих систем» содержит 74 страницы текстового документа, 22 использованных источника, 19 иллюстраций, 39 формул.

АДАПТИВНАЯ ОБУЧАЮЩАЯ СИСТЕМА, ИНФОРМАЦИОННО-ТЕРМИНОЛОГИЧЕСКИЙ БАЗИС, МОДЕЛЬ, ПРОГРАММА, ОБЪЕКТ, ИНФОРМАЦИЯ.

Объект изучения – адаптивные системы обучения иностранным языкам.

Цель работы – разработка, обоснование и реализация программно-алгоритмического комплекса, обеспечивающего формирование мультилингвистического информационно-терминологического базиса для адаптивных обучающих систем.

В результате данной работы была разработана система, позволяющая детально исследовать структуру базиса и оптимизировать разбивку словаря на модули. Кроме этого, был разработан модуль обучения на основе оптимальной разбивки словаря и выбранных методов контроля; а также в процесс обучения внедрен метод интервальных повторений.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1 Отбор языкового материала, методы изучения и запоминания иностранной лексики и компьютерные технологии обучения иностранным языкам	8
1.1 Основные подходы к отбору языкового материала.....	8
1.2 Методы изучения и запоминания иностранной лексики	14
1.2.1 Кривая забывания Эббингауза	14
1.2.2 Принцип интервальных повторений Пимслера	16
1.3 Особенности компьютера как средства обучения	19
1.4 Классификация компьютерных программ обучения.....	24
1.4.1 Классификация принципов построения обучающих систем.....	24
1.4.2 Классификации компьютерных программ для обучения иностранному языку	28
1.5 Процесс обучения как управляемый процесс	31
2 Оптимизация информационно-терминологического базиса мультилингвистической обучающей технологии.....	42
2.1 Информационно терминологический базис мультилингвистической обучающей технологии.....	42
2.2 Моделирование динамики изучения информационно-терминологического базиса с использованием цепей Маркова	43
2.3 Оптимизация структуры информационно-терминологического базиса. ...	47
2.4 Задача об оптимальной разбивке информационно-терминологического базиса на модули	54
2.5 Формирование модулей по семантическому принципу.....	55
2.6 Алгоритм случайного поиска.....	57
3 Система мультилингвистического обучения	61
3.1 Рекомендуемые требования к аппаратному обеспечению.....	61
3.2 Требования к программному обеспечению	61
3.3 Установка программы.....	61

3.4 Работа мультилингвистической обучающей системы	61
3.5 Режим «Эксперт».....	62
3.6 Режим «Обучение».....	66
ЗАКЛЮЧЕНИЕ	71
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	72

ВВЕДЕНИЕ

Широкое внедрение информационных технологий в различные сферы образовательной деятельности и наличие огромного количества разработчиков специального программного (ПО) и информационного обеспечения (ИО) компьютерных систем обучения ставит проблемы повышения эффективности разработки и функционирования этих видов обеспечения, что является одним из факторов успешного развития компьютеризации образования. В рамках методологии системного анализа и адаптации сложных систем отмечается, что внедрение в обучение новых информационных технологий следует рассматривать не только в рамках реализации учебного процесса, но и как фактор, влияющий на структуру программного и информационного обеспечения автоматизированной системы обучения, процесса его формирования, внедрения и сопровождения.

Актуальность проблемы овладения иностранным языком в современных условиях всемирной глобализации не вызывает сомнений. Количество публикаций о научных и технических достижениях растет в геометрической прогрессии. Считают, что оно удваивается, каждые пять лет. Поэтому современного квалифицированного специалиста трудно себе представить без знания иностранного языка, без умения ориентироваться в зарубежной литературе по той отрасли, в которой он работает. В связи с этим возникает вопрос об эффективности обучения языку.

В рамках решения этой проблемы обучение иностранному языку во многих неязыковых ВУЗах имеет профессиональную ориентацию. Одним из основных моментов в таком обучении является освоение характерной для учебной специальности иностранной терминологии.

Однако требование повысить уровень обучения иностранным языкам в неязыковом вузе вступает в противоречие с некоторыми условиями этого обучения, одним из которых является ограниченное количество часов, отводимых на изучение языка. Выходом из этого противоречия может стать

лишь оптимизация процесса обучения. Для этого необходимо не только найти эффективные, наиболее экономные приемы обучения языку, но и определить тот лингвистический материал, на котором можно успешно применить эти приемы.

По этим и многим другим причинам, главное требование к современным компьютерным обучающим системам состоит в том, чтобы обеспечивать максимальную степень индивидуализации процесса обучения, т.е. его адаптации к каждому конкретному ученику, что не всегда осуществимо при традиционных методах массового обучения. Данное свойство адаптивной технологии дает возможность самостоятельного обучения.

Адаптация в обучающей системе — это процесс изменения параметров и структуры модели объекта (обучаемого) и обучающих воздействий на основе текущей информации, получаемой в ходе обучения, с целью достижения оптимального состояния объекта при его начальной неопределенности и изменяющейся среде.

Следует отметить, что развитие информационных технологий на сегодняшний день позволяет разработать эффективный программно-алгоритмический аппарат для создания компьютерных адаптивно-обучающих систем, в том числе, и для изучения иностранной терминологической лексики.

Объектом исследования являются адаптивные системы обучения.

Предмет исследования - мультилингвистический информационно-терминологический базис.

Целью работы является разработка, обоснование и реализация программно-алгоритмического комплекса, обеспечивающего формирование мультилингвистического информационно-терминологического базиса для адаптивных обучающих систем.

Для достижения поставленной цели решались следующие задачи:

- анализ подходов к отбору языкового материала;
- анализ методов изучения и запоминания иностранной лексики;
- исследование принципа интервальных повторений Пимслера;

- модификация алгоритма Гинзбурга для расчета частотных характеристик различных слов;
- модификация алгоритма разбиения информационно-терминологического базиса на модули, в соответствии с семантическим принципом;
- программная реализация и тестирование построенных алгоритмов.

1 Отбор языкового материала, методы изучения и запоминания иностранной лексики и компьютерные технологии обучения иностранным языкам

1.1 Основные подходы к отбору языкового материала

Любая методика разрабатывается исходя из целей и содержания обучения, и любой метод обучения иностранным языкам всегда базируется на языковом материале.

Языковой материал составляет основу первого компонента содержания обучения, имя которому — знания. К языковому материалу примыкают и способы его употребления в речи, которые не всегда сообщаются учащимся. Если грамматические слова, формы, структуры составляют языковой материал, то правила изменения, оформления, написания и соединения слов в предложении относятся к способам его употребления. Для обучения отбирается, во-первых, тот языковой материал, который подлежит, в соответствии с программой, усвоению, а во-вторых, те способы его употребления, которые учащиеся должны знать. Впрочем, способы употребления языкового материала отбираются и в том случае, если знать их учащимся не обязательно, но обязательно использовать практически. В прямых методах многие способы употребления языкового материала усваиваются интуитивно, поэтому учащиеся их специально не изучают. Но они все равно отбираются, так как становятся объектом деятельности учащихся.

В целом языковой материал складывается из фонетического, лексического, грамматического и графического материала. Это значит, что для организованного обучения иностранному языку отбираются те звуки, буквы, слова, словосочетания, фразы, грамматические формы и структуры, которые необходимы для достижения поставленной цели обучения.

Отбор языкового материала, удовлетворяющего данные цели обучения, представляет собой сложную проблему. Ее решали и решают по-разному. Так,

В. Л. Скалкин насчитал семь подходов к отбору языкового материала, имея в виду и личный опыт обучающего (его интуицию при отборе), и отбор в границах какой-либо темы, и отбор в зависимости от языкового наполнения отобранных текстов, и выделение наиболее частотных лексических единиц и грамматических явлений, и анкетирование носителей языка и др. В. А. Томилов считает, что отбор лексического минимума (который составляет основную часть языкового материала), следуя вслед за историей развития языкознания, начинался с элементарного подхода (изучались свойства слова как элемента языковой системы), прошел структурный подход (выявлялись связи слова в речи) и переживает сейчас период системного подхода. Нам кажется, что В. А. Томилов несколько схематизирует историю методических исследований в области обучения иностранным языкам. И сегодня в отборе лексических минимумов мы сталкиваемся с критериями частотности или тематичности, выдвинутыми еще Я. А. Коменским, или с принципами семантической ценности и стилистической нейтральности.

Во всяком случае, в настоящее время существует большое количество критериев или принципов, которые используются в том или ином подходе к отбору языкового материала. Эти подходы и критерии полезно себе четко представлять. Всего, на наш взгляд, просматривается три подхода.

Во-первых, эмпирический подход. Как явствует из названия подхода, речь идет о личном опыте методиста или автора учебника. Каждый человек, знающий тот или иной язык, интуитивно может определить языковое явление, нужное для достижения целей обучения, особенно если эти цели предусматривают общение на иностранном языке. Поэтому личный опыт обычно позволяет правильно определить лексический и грамматический материал на начальной стадии обучения. Трудности в отборе языкового материала при эмпирическом подходе начинаются на более поздних стадиях обучения, когда опыт преподавателей, в том числе и авторов учебников, перестает совпадать.

К эмпирическому подходу относится и критерий отбора по языковому наполнению учебных текстов. В этом случае сначала для каждого занятия отбираются тексты, в них выделяются новые слова и грамматические формы, которые и включаются в языковой материал, подлежащий изучению. Отбор лексического и грамматического материала по текстам имеет свои положительные и отрицательные стороны. Тексты подбираются, как правило, уже на продвинутом этапе и в зависимости от темы, определяемой программой. Это значит, что они должны содержать тематическую лексику и типичные для данного стиля речи грамматические явления. Все это способствует отбору требуемого языкового материала. С другой стороны, тексты могут содержать и случайную лексику, а также весьма редкие грамматические явления. Их отбор в качестве языкового материала урока, на котором учебные тексты прорабатываются, ничем не оправдан. Значение тех или иных «случайных» лексических единиц можно определить по словарю, а значение грамматических форм — по контексту. Недаром И. М. Берман вводит для отбора грамматических явлений при обучении чтению критерий прозрачности, исключающий все те явления, которые могут быть поняты по аналогии, на основе соотношения уже известных элементов или контекста.

Как видно из сказанного, полностью отрицать эмпирический подход нет оснований. Он неплохо обеспечивает языковым материалом начальный период обучения и способствует выделению тематической лексики.

Во-вторых, следует иметь в виду лингвистический подход. Язык составляет предмет обучения, а потому тесная связь всех проблем обучения с языком вполне естественна. Лингвистические характеристики языкового материала дают методике большое количество критериев отбора. Назовем наиболее распространенные из этих критериев (или принципов): критерий сочетаемости слова, критерий многозначности слова, критерий семантической ценности, критерий стилистической нейтральности, критерий словообразовательной ценности, критерий строевой способности и др.

Определенный интерес представляет критерий (принцип) сочетаемости. Имеется в виду способность слова сочетаться в речи с другими словами. Причем речь идет не только о компонентах сложных слов (например, *pomme de terre*), но и о сочетаниях типа *manger une pomme*, *une pomme au four*, *une petite pomme* и т. д. Чем больше сочетаний получается из отобранной лексики, тем шире лексическая база учащихся.

Критерий многозначности предполагает выделение, в первую очередь, тех языковых знаков, которые способны обозначать несколько объектов действительности. Так, глагол *charger* имеет до 10 значений, и его включение в базовую лексику представляет определенный выигрыш. Это не значит, что все значения необходимо сообщать учащимся, но знание этого глагола значительно расширит их потенциальный словарь.

Критерий семантической ценности заключается в выделении наиболее важных понятий из окружения школьников, жизни родной страны и страны изучаемого языка.

Критерий стилистической нейтральности говорит о необходимости исключать из базового словаря диалектизмы, профессионализмы, архаизмы и все то, что остается за пределами нейтрального стиля.

Критерий словообразовательной ценности предполагает включение в отобранный словарь слов, способных давать значительное число производных лексических единиц, служащих основанием для расширения потенциального словаря. Так, например, от существительного *fin* легко перейти к глаголам *finir*, *définir*, прилагательным *final*, *fin* и др.

Критерий строевой способности требует включать в языковой минимум предлоги, артикли, союзы, вспомогательные глаголы, модальные слова.

Как явствует из сказанного, лингвистический подход к отбору языкового материала включает семантические, грамматические и стилистические критерии. Эти критерии дают определенный методический выигрыш, однако возникающие в ряде случаев противоречия (семантическая ценность слова, не

имеющего словообразовательной ценности) не всегда позволяют их использовать достаточно эффективно.

Третий подход можно назвать прагматическим. Он исходит не из лингвистических свойств элементов языка, а из требований общения на иностранном языке. Эти требования породили следующие критерии (принципы) отбора: критерий частотности, тематический критерий, или критерий наличности, критерий употребительности, критерий описания понятий, критерий исключения синонимов, критерий прозрачности и др.

Чаще всего в основу отбора языкового минимума кладется критерий частотности. Действительно, чем чаще слово или грамматическая структура встречаются в речи, тем они по-видимому, нужнее. Кроме того, критерий частотности предполагает статистическую обработку устных и письменных текстов, т. е. он обретает научную достоверность. Критерий частотности оказался одним из двух принципов, положенных в основу известного словаря Г. Гугенейма «*Dictionnaire fondamental de la langue française*», содержащего примерно 3000 слов. Опыт составления словаря основной лексики французского языка показал, что наиболее частотными лексическими единицами являются служебные слова, затем идут глаголы, прилагательные и только потом существительные, причем общего характера. Конкретные существительные принцип частотности почти не выделяет. Кроме того, критерий частотности позволяет выделить только первые несколько сотен слов. Далее проблема отбора остается, так как частотность слов становится примерно одинаковой. Это обстоятельство заставило составителей словарей прибегнуть еще к одному критерию — тематичности (наличности).

Тематический критерий обычно понимается как отбор слов заданной тематики. Составители базового словаря французского языка, расширив его «компетентность», получили критерий наличности. Этот критерий им понадобился для того, чтобы пополнить лексику, отобранную по принципу частотности и не содержащую конкретных существительных. Они действовали следующим образом: выбрали 16 тем (например, части тела, одежда, продукты

питания) и предложили ученикам лицеев написать для каждой темы по 20 наиболее важных с их точки зрения слов. Слова (в основном существительные), записанные большинством школьников, и составляли по каждой теме так называемый наличный словарь (*vocabulaire disponible*). Эти слова, по мнению руководителей эксперимента, появляются в сознании всякий раз, когда в них возникает необходимость. Так, в базовом словаре французского языка появились нечастотные, но очень нужные слова *fourchette* — вилка, *nez* — нос, *bouton* — пуговица. Как видно из сказанного, тематический критерий очень удачно дополняет критерий частотности.

Критерий употребительности суммарно учитывает основные лингвистические характеристики лексических единиц, а именно: семантическую ценность, стилистическую нейтральность, сочетаемость, строевую способность, словообразовательную ценность и многозначность. Учет всех этих характеристик позволяет вывести коэффициент суммарной употребительности каждой лексической единицы и соответственно решать проблему отбора.

Критерий описания понятий отдает предпочтение тем словам, с помощью которых можно описать значение слова, а также словам, которые могут служить синонимической заменой других слов. Так, слово *élève* имеет преимущество перед словами *écolier*, *lycéen*, *collégien*, *étudiant*, *normalien*, *disciple* и др., так как может их заменять или войти в формулировку, определяющую понятия, обозначаемые этими словами (например: *normalien* — *l'élève de l'école normale*).

Критерий исключения синонимов предлагает оставлять для словаря в ряду синонимов только наиболее употребительное и нейтральное в стилистическом отношении слово.

О критерии прозрачности, используемом И. М. Берманом для отбора грамматических понятий, уже говорилось.

Критериев или принципов отбора значительно больше, чем было представлено в этой работе. Их обилие говорит о продолжающихся поисках

оптимального подхода к отбору языкового материала, в которых четко прослеживается три пути: эмпирический, лингвистический и прагматический. Но наиболее применительной для данной адаптивной-обучающей системы является критерий семантической ценности (лингвистический подход).

1.2 Методы изучения и запоминания иностранной лексики

Основой успеха в изучении любого иностранного языка является обширный словарный запас, который поможет свободно общаться и выражать свои мысли. Запоминание новой лексики является основной проблемой, с которой сталкиваются многие люди, изучающие иностранный язык. Следует помнить, что важно не только выучить новые слова, но и регулярно увеличивать их количество.

Таким образом, одно из основных условий хорошего владения иностранным языком - большой словарный запас. Хорошим уровнем знания языка считается освоение более 15 тысяч слов, а владение 50 тысячами слов считается уже отличным уровнем.

Существует огромное количество различных методик для изучения иностранных языков, в том числе предназначенных для запоминания новых слов. Одним из самых эффективных способов пополнения словарного запаса является метод интервальных повторений.

1.2.1 Кривая забывания Эббингауза

Первооткрывателем метода «повторений» является немецкий психолог Герман Эббингауз (Hermann Ebbinghaus).

Для своих опытов Эббингауз подобрал бессмысленные слоги, состоящие из двух согласных и гласной между ними, и постарался чтобы эти слоги не вызывали никаких мысленных ассоциаций.

Ученый обнаружил, что сразу после первого запоминания, т.е. момента, когда испытуемый безошибочно повторил серию таких слогов, начинается процесс забывания. Вначале процесс идет очень быстро, и спустя один час люди забывают до 60% информации. Через 10 часов — 65%, и далее все медленнее. Через шесть дней процесс приостанавливается на значении 20%.

На основе кривой забывания (рис. 1) Эббингауз подготовил методику под названием «Запомнить надолго». Методика подходит для любого типа данных - от бессмысленных слогов до произведений величайших писателей. От точных наук, до естествознания.



Рисунок 1 – Кривая забывания Эббингауза

В результате исследований были сделаны выводы:

- Через 20 минут забывается 40% заученной информации;
- Через 60 минут - 50%;
- Через день - 70%;
- Через месяц – 99,9%.

График (рис. 2) демонстрирует процесс изучения иностранных слов без использования метода интервальных повторений. Мы видим, что методика

«зазубривания» (в данном случае, попытки выучить 150 слов в один подход) неэффективна. Вывод: большая часть информации теряется в первые часы.

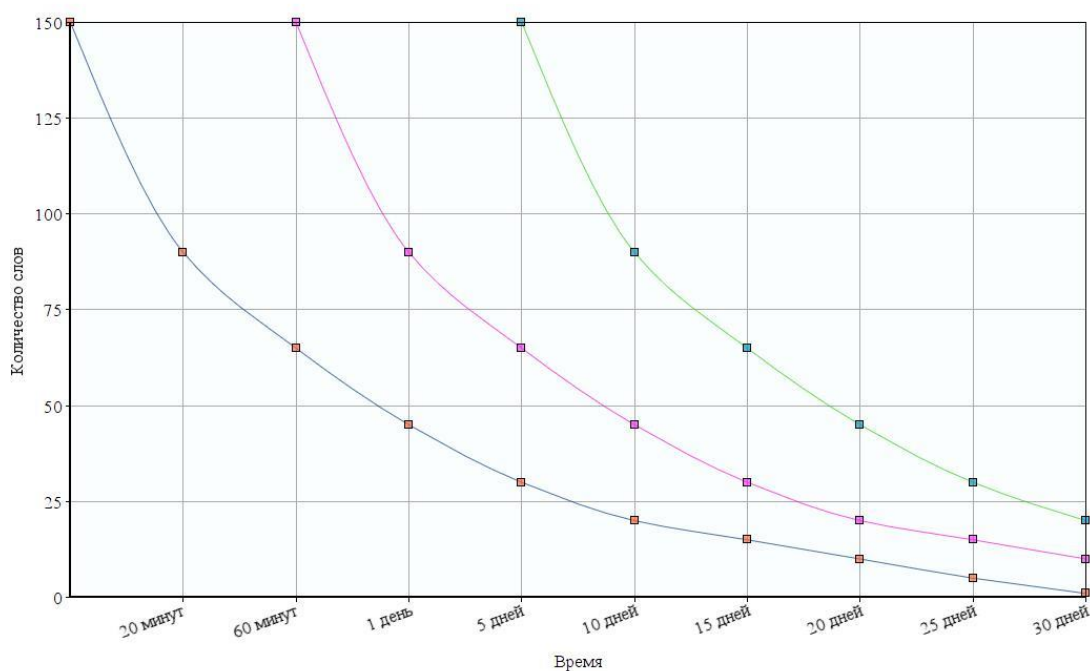


Рисунок 2 – Изучение иностранных слов без повторений.

1.2.2. Принцип интервальных повторений Пимслера

Альтернативный подход предложил в 1967 году американский лингвист Пол Пимслер (Paul Pimsleur), много лет посвятивший вопросу изучения языков. Так же, как и Эббингауз, он заметил, что, если пытаться учить новые слова, вероятность забыть его моментально огромная. Но если же повторять слово как раз перед тем, как оно окончательно стёрлось из памяти, время, через которое человек снова начнёт его забывать, увеличится экспоненциально. То есть, если бы вы забывали новое слово через 5 секунд, повторив его через этот промежуток времени, вы укрепили бы его в памяти на 25 секунд, повторив слово после этих 25 секунд, вы укрепили бы его на 10 минут и т.д.

Уникальной данный метод делают 4 принципа, на которых строится обучение любому иностранному языку.

1. Градуированные интервальные повторения

Этот принцип заключается в повторении изученных слов через увеличивающиеся интервалы и является вариантом метода запоминания с помощью интервальных повторений. Например, обучаемый повторяет новое слово сначала через несколько секунд, затем — через несколько минут, через несколько часов и через несколько дней. Цель интервальных повторений — «записать» изучаемые слова в долговременную память обучаемого.

2. Антиципация

Традиционное обучение иностранному языку подразумевает многократное пассивное повторение за инструктором. Но по мнению Пимслера этого недостаточно. Его методика побуждает обучаемых участвовать в самом процессе обучения. Для этого он разработал технику «стимула и реакции», когда ученикам предлагается предугадать ответ на заданный вопрос, что заставляет их активно вспоминать только что заученные слова. Пимслер утверждал, что принцип антиципации (предвосхищения) воплощается в естественном общении, когда говорящий должен быстро строить фразы.

3. Естественное обучение

Система Пимслера исходит из того, что восприятие речи на слух — это самый естественный способ начать изучать новый язык, поскольку именно так мы в детстве начинаем учиться родному языку. Потом человек начинает говорить и только потом писать или читать. Поэтому его система обучения использует только аудиоформат, в которой параллельно происходит изучение грамматики, слов и произношения.

4. Словарное ядро

Метод Пимслера делает акцент на обучении наиболее часто-употребляемым словам для формирования «словарного ядра», что позволяет значительно сократить время обучения. Анализ частотности употребления слов в тексте показал, что относительно небольшой набор слов покрывает

большинство словоупотреблений в любом языке. Например, для понимания 80% англоязычных текстов достаточно усвоить 2 тыс. наиболее частотных слов. Грамматический материал никогда не преподносится напрямую, а дается через частотное повторение фраз, содержащих нужную грамматическую конструкцию. Именно так, согласно теории Пимслера, носители языка в детстве интуитивно усваивают грамматику.

Рассмотрим более подробно 1 принцип – интервальные повторения.

В основе этого метода лежит предположение о том, что существует некий идеальный момент для повторения выученной ранее информации. Если делать это раньше, чем следует, то это пустая трата времени. Если же делать это слишком поздно – вероятнее всего, мы забудем материал и, сами того не осознавая, начнем учить все заново. Идеальный момент для повторения – это время, когда вы вот-вот забудете эту информацию. Задача компьютерной программы заключается в том, чтобы рассчитать этот момент и вовремя напомнить о необходимости повторить изученное.

С каждым своевременным напоминанием, знание сохраняется в памяти на больший промежуток времени, т.е. понадобится больше времени, чтобы забыть то, что уже выучили. Следовательно, интервалы между занятиями/повторениями увеличиваются. Таким образом, программа фиксирует временную составляющую обучения и назначает следующий сеанс соответственно.

Для достижения положительных результатов при изучении новой информации, необходимо производить:

- первый повтор - сразу по окончании чтения;
- второй повтор - спустя 20 минут после первого повторения;
- третий повтор - спустя 8 часов после второго;
- четвертый повтор - через 24 часа после третьего.

На графике (рис. 3) изображен процесс изучения с использованием данной технологии. Вывод: при соблюдении всех условий, запоминается около 90 % информации.

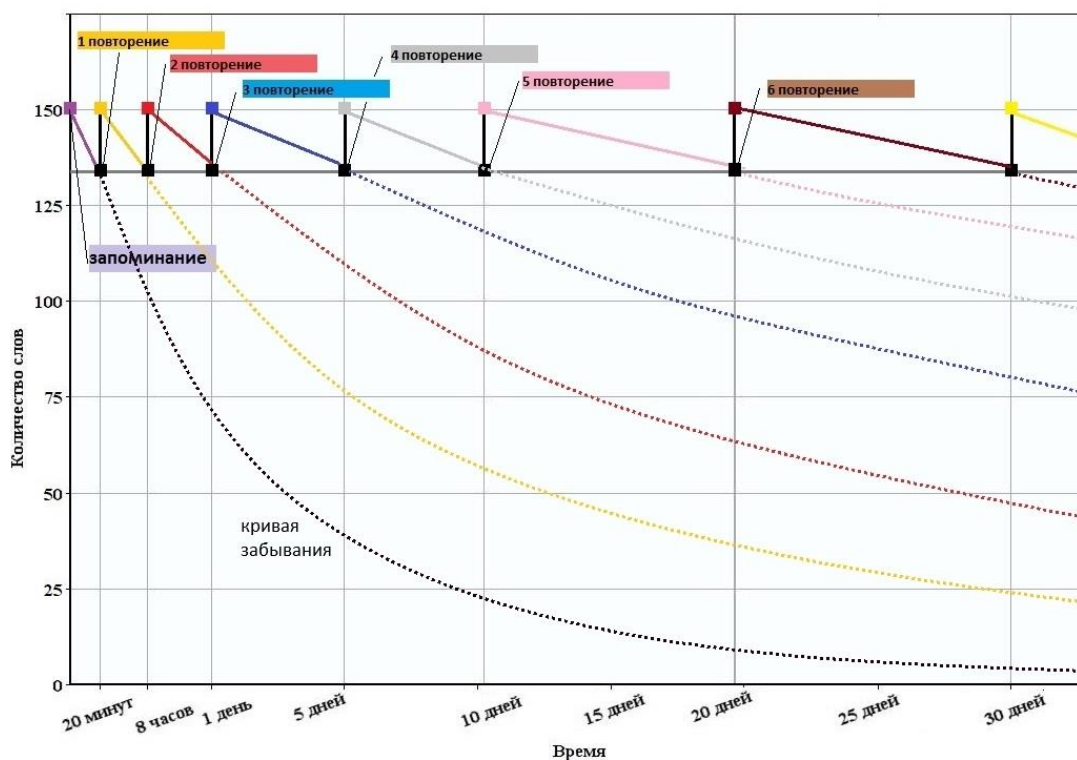


Рисунок 3 – Изучение иностранных слов с использованием метода интервальных повторений.

Метод интервального повторения подтвержден научно и является одним из самых эффективных способов пополнения словарного запаса.

1.3 Особенности компьютера как средства обучения

Задачи совершенствования учебного процесса, повышения его эффективности не новы для высшей школы. Попытки оптимизации учебного процесса предпринимаются, главным образом, в трех направлениях [6]:

- разработка новых методов и приемов обучения;
- внедрение новых форм организации учебного процесса (основной упор в последнее время делается на повышение роли самостоятельной работы обучающихся во внеаудиторное время);
- внедрение в практику преподавания различных видов информационных технологий.

Традиционная - фронтальная - форма обучения порождает ряд трудно разрешимых противоречий, связанных:

- с возрастанием объемов учебной информации, подлежащей усвоению и контролю, при сохранении или даже сокращении количества аудиторных часов;
- с дифференциацией обучения (по различным параметрам) в самой учебной группе.

Таким образом, важнейшими направлениями в решении отмеченных задач являются:

- а) интенсификация;
- б) индивидуализация обучения.

В данном случае многие надежды педагогов связаны с внедрением различных технических средств обучения, технических средств информации и технических средств контроля.

Применение технических средств информации и технических средств контроля дало положительные результаты. Технические средства информации (средства статичной и динамической проекции) позволили повысить доступность изложения материала, демонстрировать различные процессы и тем самым существенно уплотнить занятие, насытить его наглядностью.

Технические средства контроля призваны осуществлять систематический контроль студентов по уровню усвоения всего объема учебного материала.

Технические средства обучения могут быть использованы как в структуре аудиторного занятия, так и в условиях внеаудиторной самостоятельной работы.

В настоящее время решены технические проблемы создания программ обучения, обеспечивающие:

- а) адаптивность учебного материала к уровню подготовленности (в зависимости от индивидуальных особенностей студентов);
- б) многотерминальность (одновременная работа группы пользователей);
- в) интерактивность (взаимодействие компьютера и студента, имитирующее в известной степени естественное общение);

г) подконтрольность индивидуальной работы студентов во внеаудиторное время.

В последнее время открылись новые возможности и перспективы, в связи с появлением персональной компьютерной техники и внедрением ее в учебный процесс.

Компьютер – это особый вид технических средств обучения. В силу своих технических возможностей современный компьютер представляется наиболее подходящим техническим средством обучения для организации самостоятельной работы студентов как во внеаудиторное время, так и во время аудиторного занятия.

Компьютеры способны во многом решать те же методические задачи, что и традиционные технические средства обучения. Но в условиях компьютерного обучения это делается на более мощной, совершенной и быстродействующей технике. Компьютер реализует обучение в диалоговом (компьютер студент) режиме.

Компьютеризованные учебные материалы (учебные компьютерные программы) способны полнее и глубже адаптироваться к индивидуальным особенностям студентов.

Обусловлено это спецификой компьютера как одного из видов технических средств обучения, которая состоит в следующем:

1. Значительный объём памяти современных компьютеров, что позволяет хранить и оперативно использовать большие объёмы учебной информации (формулировки заданий, тексты, упражнения, примеры и образцы, справочную - корректирующую и консультирующую - информацию, разнообразные ремарки – реакции на те или иные действия студента и т.п.).

2. Высокое быстродействие компьютера (сотни тысяч операций в секунду). Это позволяет значительно повысить реактивность данного вида информационных технологий.

3. Возможность не только предъявлять учебный материал, фиксировать ответы (это могут делать магнитофон и обучающие машины), но и

анализировать ответы и запросы студентов, а это самое главное для самостоятельной работы.

Анализ ответов и вопросов обеспечивается дидактическими возможностями компьютерных программ обучения и контроля [6].

1. Реализация такой связи учебного материала (компьютерной программы) с обучаемым, которая ведется в диалоговом режиме, имитируя некоторые функции преподавателя и, в известной степени, общение.

2. Наличие обратной связи, под которой мы понимаем возможность осуществления коррекции ответа самим обучаемым с опорой на консультирующую информацию. Консультирующая информация выбирается из памяти компьютера либо самим студентом, либо на основе автоматической диагностики ошибок, допускаемых студентом в ходе работы. Способ предъявления подобного рода информации зависит от типа учебной компьютерной программы.

3. Из возможности диагностики ошибок вытекает и такая особенность компьютера (точнее, компьютерной программы), как адаптивность. Компьютеризированный урок проходит с учетом индивидуальных особенностей студентов. Изучение, тренировка, повторение и контроль одного и того же материала может осуществляться:

- с различной степенью глубины и полноты;
- в индивидуальном темпе;
- в индивидуальной последовательности (выбираемой самим студентом).

4. Возможность проводить в автоматическом режиме многофакторный сбор и анализ статистической информации, получаемой в процессе компьютеризованного занятия, без нарушения естественности протекания компьютерного урока. При этом компьютер способен фиксировать достаточно большое количество параметров:

- время, затраченное студентами на работу со всей программой, группой заданий или с каким-либо конкретным заданием или упражнением;

- количество верных/неверных ответов (причем, неверные ответы могут «раскладываться по полочкам» в памяти компьютера);
- количество обращений к справочной информации, а также характер наиболее часто запрашиваемой помощи той или иной группой обучаемых;
- количество попыток при выполнении заданий. Компьютер может фиксировать информацию и по другим параметрам. Это определяется авторами учебных программ в зависимости от целей и задач, выдвигаемых педагогами.

Собранная и обработанная компьютером статистическая информация может быть преобразована в так называемые граф – схемы движения по программе отдельных студентов или учебных групп (в усредненном виде) в различные временные отрезки. Кроме того, указанная в пункте 7 возможность компьютера, как одного из средств информационных технологий, предоставляет студенту и преподавателю в обобщенном и обработанном виде объективную информацию о динамике обучения.

Тем самым компьютер выполняет исключительно важную миссию через подсистему статистического анализа процесса обучения, неосуществимую в условиях самостоятельной работы студентов никакими иными техническими средствами обучения, т. е. помогает:

- студенту внести коррективы в тактику самообучения;
- преподавателю выработать индивидуальный подход как к отдельному обучаемому, так и к группе в целом.

Таким образом, внедрение компьютеров в учебный процесс придает новый статус самостоятельной работе студентов, в рамках которой обучение с использованием компьютерной техники, являясь по форме индивидуальным и самостоятельным, становится по сути контролируемым и управляемым.

1.4 Классификация компьютерных программ обучения

Центральное место в программном обеспечении обучения занимают обучающие программы. Существует огромное разнообразие обучающих программ и их классификаций.

1.4.1 Классификация принципов построения обучающих систем

Каждая автоматизированная обучающая система имеет определенную структуру на основе группы элементов с указанием связей между ними и дающее представление о системе в целом. Поэтому структура системы может быть охарактеризована по имеющимся в ней типам связей.

В [5] представлена следующая классификация автоматизированных обучающих систем (рис. 4):

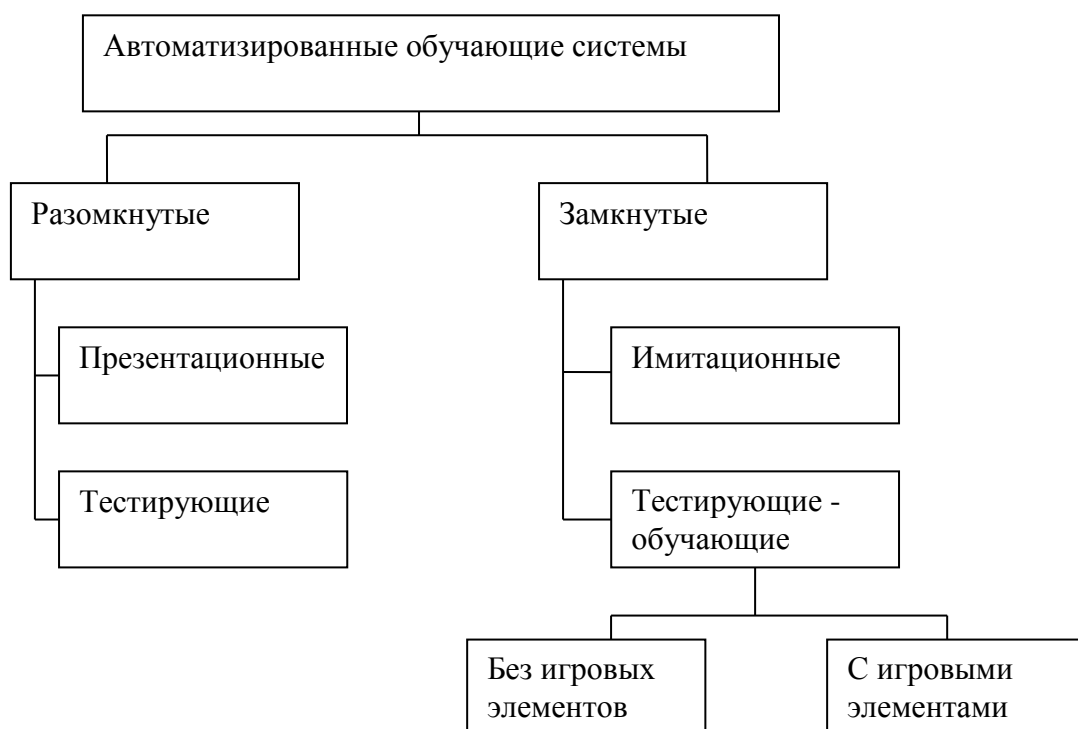


Рисунок 4 – Классификация АОС

По структурным признакам взаимодействия обучающей системы с пользователем автоматизированной обучающей системы подразделяются на два базовых класса: разомкнутые (без обратной связи) и замкнутые (с обратной связью) системы, которые отличаются принципиальным подходом к процессу обучения.

В разомкнутых автоматизированных обучающих системах не учитываются отклики учащихся на поставленные вопросы и не корректируется последовательность предъявления учебного материала в функции степени усвоения учащимся изучаемой темы.

В автоматизированных обучающих системах данного типа присутствует только прямая информационная связь между системой и учащимся, которому последовательно предоставляется визуальная информация. При этом обучаемый находится в режиме пассивного наблюдателя, от которого не требуется ни каких откликов по взаимодействию с автоматизированной обучающей системой.

В тестирующих автоматизированных обучающих системах без обратной связи основной упор делается на выявление уровня знаний учащихся в определенный период учебного процесса. Используя различную методику, такие системы предъявляют обучаемому открытый или закрытый вариант вопроса (вопрос с вариантами выбора ответа). От учащегося ожидается отклик в виде ответа на поставленный вопрос. Ответ фиксируется в блоке фиксатора ошибок. По результатам опроса выставляется определенный балл, который служит критерием для результирующей оценки по степени усвоения учащимся требуемого учебного материала.

Наиболее широкими функциональными возможностями и высокой эффективностью в учебном процессе обладают автоматизированные обучающие системы, где организована обратная связь между учащимся и обучающей системой. Подробнее процесс обучения как управляемый процесс рассмотрен в главе 1.3.

При реализации любой из ранее рассмотренных структур АОС используются вполне определенные алгоритмические подходы, диктуемые методикой проведения учебного занятия. Обычно любая обучающая система представляет собой совокупность порций информации, которая в той или иной форме предъявляется ученику. Современная вычислительная техника обладает широкими функциональными возможностями и позволяет использовать информацию, представленную в виде обычного текста, графического изображения, аудио и видео фрагментов. При этом можно сосредоточить все средства представления информации, существующие в настоящее время для повышения эффективности учебного процесса. В настоящее время, как наиболее оптимальный вариант в большинстве курсов программированного обучения, применяют текстовое и графическое представление информации.

Классификация обучающих систем по алгоритмическому построению представлена на рисунке 5:

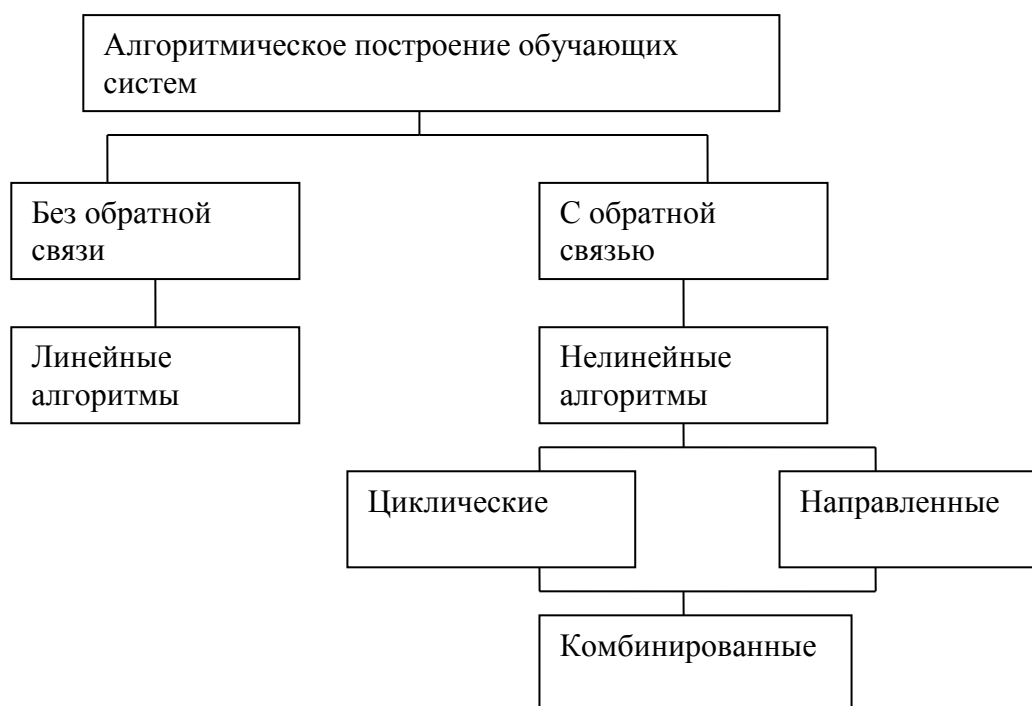


Рисунок 5 - Классификация обучающих систем по алгоритмическому построению

При использовании линейных алгоритмов в обучающих системах учащемуся, согласно методики, последовательно предъявляется информация. В

качестве достоинств линейного алгоритма в обучающих системах можно отметить простоту разработки такой системы, а в качестве недостатков - трудоемкость раскрытия некоторых тем и невозможность гарантированного закрепления полученных знаний. В обучающих системах построенных с использованием нелинейных алгоритмов появляется возможность изменять последовательность предъявления информации в зависимости от того или иного отклика учащегося на информационное воздействие.

Нелинейные алгоритмы, в свою очередь, делятся на циклические, направленные и комбинированные.

Циклические алгоритмы предполагают повторный возврат к информации, отражающим темы, которые учащийся недостаточно усвоил. Если ученик принимает неверное решение поставленной ему задачи, то обучающая система может повторно предъявить информацию, которая уже ранее была показана, для повторного прохождения темы или ее закрепления.

Направленные алгоритмы в зависимости от принятого решения учащимся выбирается та или иная последовательность представления информации и возврата обратно не предполагается. В комбинированных алгоритмах используется оба принципа. По результатам решения принятого учащимся для ответа на вопрос, обучающая система изменяет последовательность предъявления информации, однако на N-ом шаге возможен возврат к предыдущей порции информации.

Таким образом, каждая из представленных структур позволяет предъявлять учебный материал, в соответствии с последовательностью, который обеспечивает приемлемое предъявление учебного материала в соответствии с требованием предметной области.

1.4.2 Классификации компьютерных программ для обучения иностранному языку

Рассматривая обучающие программы в целом, независимо от изучаемой дисциплины, ряд авторов использует в качестве оснований для классификации следующие критерии:

- цели обучения;
- возможность интеграции программ в учебный процесс;
- степень самостоятельности / активности обучаемого.

В соответствии с этими критериями все компьютерные обучающие программы подразделяются на четыре основных типа:

1. Тренировочно-закрепительные программы для обучения отдельным навыкам; такие программы легче всего интегрируются в учебный процесс и совмещаются с любой стратегией обучения;

2. «Тьюторские» программы, включающие материал и виды заданий, которым преподаватель не может уделять много времени в классе, и позволяющие индивидуализировать процесс обучения; они эффективны в качестве дополнения к аудиторному обучению.

3. Моделирующие программы, предоставляющие уникальные возможности исследований, имитации различных процессов; они дают возможность обучаться в ходе самостоятельной деятельности. Недостатком таких программ является то, что бывает трудно проверить, чему конкретно научился ученик в процессе работы с программой;

4. Учебные игры – сходны с моделирующими программами и отличаются от них привнесением элемента соперничества.

В исследованиях, классифицирующих программы для обучения языку, наряду с представленными общими основаниями учитываются основания, применимые только к компьютерным программам обучения данной дисциплине. Ориентация на методическое назначение программы позволяет выделить следующие их:

- для изучения различных аспектов языка: фонетики, лексики, грамматики;
- позволяющие формировать навыки в различных видах речевой деятельности: чтении, письме, говорении, аудировании;
- предназначенные для развития языковой, речевой и коммуникативной компетенции обучаемых;
- лингвострановедческие;
- игровые программы.

Ряд авторов рассматривает обучающие программы на основе соотношения их содержания с существующими учебными курсами. В этом случае программы подразделяются на автономные (независимые от учебного курса) и локальные (являющиеся дополнением к конкретному учебному пособию).

Приведенные выше классификации дополняются такими существенными показателями, как структура программы и тип обратной связи. В зависимости от наличия в структуре таких этапов изучения материала, как инструктаж, презентация, тренировка; коррекция, дополнительные презентация и тренировка; контроль, программы подразделяются на:

- контролирующие (содержащие только этап контроля);
- обучающие (содержащие все этапы);
- контрольно-корректировочные (этапы коррекции, дополнительной презентации, тренировки и контроля).

Обратная связь может быть представлена:

- только оценкой правильности / неправильности ответа («drill and practice»);
- разветвленной системой помощи и включением серии дополнительных корректировочных упражнений и разнообразных тестов («tutor»).

В работе [2] предлагается классификация компьютерных обучающих программ на основе соотношения особенностей лингвистического материала и способов решения учебных заданий и выделяется пять типов программ:

- построенные на заданиях, требующих установления соответствия между словом и его значением;
- включающие грамматические упражнения;
- ориентированные на решение коммуникативных задач в различных ситуациях общения;
- реконструкции текста;
- создания текста, предназначенные для формирования навыков подготовки текстов заданной структуры – аннотаций, рефератов, рецензий, писем и т.д.

Классификация обучающих программ по типам заданий, существующая в зарубежной теории компьютерного обучения языку, включает следующие их виды:

- заполнение пропусков (gapfilling);
- множественный выбор (multiple choice);
- восстановление порядка слов в предложении, последовательности предложений (sequencing);
- установление соответствий (matching);
- реконструкция текста (total-deletion).

Необходимо отметить, что современные программы расширяют рамки видов речевой деятельности и аспектов языка, которым можно обучать с помощью компьютера.

Рассмотренные классификации компьютерных программ для обучения иностранному языку показывают, что современные программы позволяют обучать всем видам речевой деятельности и различным аспекта языка, включать разнообразные виды упражнений и формы представления обучающей информации, таким образом, повышая степень индивидуализации процесса обучения.

1.5 Процесс обучения как управляемый процесс

Компьютерные технологии обучения обеспечивают адаптацию процесса обучения к индивидуальным характеристикам обучаемых, разгружают преподавателей от ряда трудоемких и часто повторяющихся операций по представлению учебной информации и контролю знаний, способствуют разработке объективных методов контроля знаний и облегчают накопление передового учебно-методического опыта. При проведении занятий в классах, оснащенных компьютерными технологиями обучения, возрастают активность обучаемых, самостоятельно прорабатывающих большой объем учебной информации, и возможности преподавателя по управлению учебной деятельностью группы обучаемых [3].

Цели обучения. В психологии целью называют осознанный образ или словесное представление будущего результата действия. В кибернетике – формальное описание конечной ситуации, к достижению которой стремится целенаправленная система. Под целью обучения понимается описание состояния знаний, умений и других характеристик обучаемого, которое должно быть достигнуто в результате его работы с компьютерными обучающими программами. Качество управления определяется временем, затраченным на перевод обучаемого в заданное состояние. Реальное описание целей содержит неопределенные параметры, значения которых уточняются в процессе обучения.

Методы обучения. В самом общем виде в системе обучения взаимодействуют три основных элемента: преподаватель, обучаемый (студент) и учебный материал. Ее функционирование обеспечивается двумя субъектами целенаправленной деятельности. Деятельность преподавателя (преподавание) заключается в передаче знаний, демонстрации умений, организации деятельности студента при изучении дисциплины. Деятельность студента (учение) состоит в усвоении знаний, в закреплении и применении умений по решению учебных задач.

Таким образом, обучение – это совместная деятельность преподавателя и студента по организации способа обмена знаниями, познавательной и практической деятельности. Естественно, что методы (общие правила организации) обучения включают в себя комбинации методов преподавания и учения. Методы преподавания определяют правила организации и предъявления учебного материала, способы управления учебной деятельностью студентов с целью эффективной передачи необходимых знаний. Методы изучения – правила самостоятельной деятельности, направленной на усвоение новых знаний и умений.

С точки зрения теории управления методы обучения отличаются степенью участия обучаемого в процессе управления обучением. В первой группе методов предполагается, что преподаватель излагает учебный материал, демонстрирует операции с учебным материалом и управляет процессом усвоения его по заранее составленной жесткой схеме. Обучаемый при таком способе внешне пассивен. Степень усвоения определяется при итоговом или промежуточном контроле [3].

Второй группе методов соответствует более активное поведение обучаемых, которые отвечают на текущие вопросы преподавателя и имеют возможность сами задавать вопросы, уточняя неясные выражения и операции преподавателя. При этом обеспечивается текущий контроль степени усвоения пройденного материала и корректировка схемы изучения предмета [3].

В методах третьей группы основная работа преподавателя заключается в организации условий для активной самостоятельной деятельности обучаемых. Контроль усвоения ведется путем наблюдения за действиями обучаемых [3].

При более детальном членении методов обучения учитываются соответствия между методами преподавания и учения.

К числу основных методов преподавания относят:

1. Объяснительно-иллюстративный.
2. Информационно-сообщающий.
3. Инструктирующий.

4. Побуждающий.

Среди методов учения выделяют:

1. Продуктивно-практический.
2. Репродуктивный.
3. Эвристический (частично-поисковый).
4. Исследовательский (поисковый).

При разработке теории управления познавательной деятельностью нельзя становиться на упрощенную точку зрения, что техническое средство само управляет сложными психологическими процессами, одним из которых является обучение человека. Компьютерная обучающая программа только средство и является посредником между преподавателем и обучаемым: управление познавательной деятельностью происходит только в пределах модели, избранной преподавателем на этапах составления учебного курса и проведения занятий. Основным отличием системы управления познавательной деятельностью от управления техническими системами является то, что управление деятельностью человека (объект управления) осуществляется, опосредовано, через его психическую деятельность.

В соответствии с этим можно определить управление познавательной деятельностью как способ организации процесса обучения, обеспечивающий достижение дидактических целей, основными из которых являются знания, умения, навыки.

В общем случае процесс управления деятельностью включает в себя этапы [3]:

- формулирование задач управления;
- описание и анализ начального состояния объекта управления;
- выбор и определение характера управляющих воздействий на объект управления;
- реализацию управляющих воздействий (исполнительными устройствами);
- оценку результатов управления (измерение);

– введение корректирующих управляющих воздействий.

Управление процессом обучения субъекта происходит по принципу обратной связи. Учебные задания, разъяснения вырабатываются преподавателем, исходя из цели обучения, на основе сравнения модели требуемых знаний и модели текущих знаний обучаемого. Модель текущих знаний обучаемого формируется в сознании преподавателя на основе контроля результатов выполнения обучаемым учебных заданий путем сравнения их с правильными результатами [4]. Последние должны быть получены преподавателем по им же выработанному заданию.

Управление обучением включает в себя два взаимосвязанных процесса: организацию деятельности обучаемого и контроль этой деятельности. Эти процессы непрерывно взаимодействуют: результат контроля влияет на содержание управляющих воздействий, т.е. на дальнейшую организацию деятельности. В свою очередь организация определенной деятельности требует и определенную форму контроля, и конкретный способ регистрации этой деятельности. Возможны различные сочетания этих процессов и переходы от одного к другому.

Таким образом, контроль есть способ оценки не только достигнутых знаний, но и способ организации обратных связей в процессе обучения, что является необходимым условием достижения цели обучения. Обратные связи адаптируют процесс обучения к уровню освоения учебного материала, возвращая, когда это необходимо, обучаемого на предшествующие этапы обучения.

Процесс обучения, как управляемый по принципу обратной связи процесс, состоящий из операций, носит дуальный характер: в процессе обучения формируется не только его выход – знания обучаемого, но и реализация самого процесса обучения, которая в конечном итоге определяется предъявляемыми требованиями к знаниям обучаемого, их начальным состоянием и способностью обучаемого [4].

Алгоритмы управления обучением интерпретируются как алгоритмы выработки учебных заданий, разъяснений. Они должны определяться субъективными законами познания и зависят от используемых технологий, методов, приемов обучения.

Процесс обучения можно изобразить с помощью схемы, изображенной на рисунке 6. Изображенная на рисунке система обучения идентична системе управления. Здесь объект управления является объектом обучения (назовем его условно “ученик”), а управляющее устройство - обучающим устройством (назовем его условно “учитель”). “Учитель” подает на вход “ученика” порцию обучающей информации U , Y – состояние “ученика”, подаваемое на датчик (в роли датчика реально может выступать, например, контрольная работа или опрос на семинаре, т.е. тест), Y' – информация о состоянии “ученика”, получаемая “учителем”. “Учителю” сообщаются цели обучения Z^* и ресурсы R , которыми он располагает для обучения.

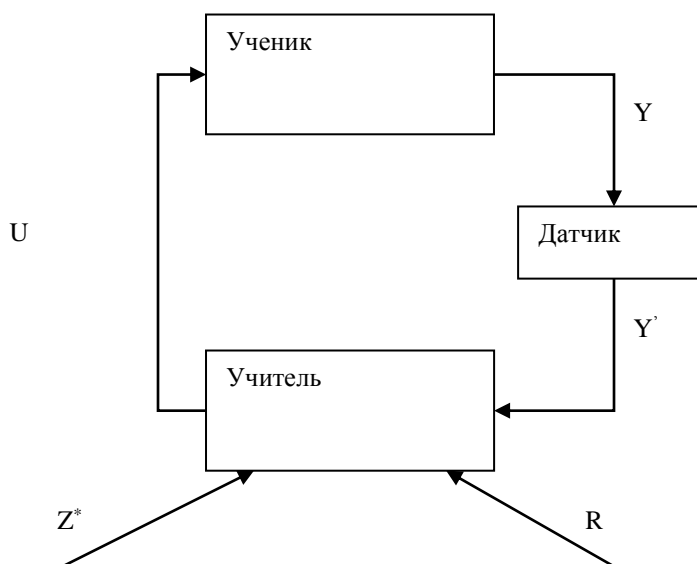


Рисунок 6 – Схема процесса обучения

Задача состоит в следующем: организовать обучение U , изменяющее состояние Y ученика таким образом, чтобы выполнялись поставленные цели обучения Z^* :

$$U=\varphi(Y', Z^*, R), \quad (1)$$

где φ - алгоритм обучения.

Так как ученик всегда является сложным объектом, точной априорной модели которого не существует, то, как известно из теории управления, без адаптации невозможно построение эффективного управления данным объектом. Главное требование к современным компьютерным обучающим системам состоит в том, чтобы обеспечивать максимальную степень индивидуализации процесса обучения, т.е. его адаптации к каждому конкретному ученику, что не всегда осуществимо при традиционных методах массового обучения. Под адаптацией в теории управления понимают «процесс изменения параметров и структуры системы, а возможно, и управляющих воздействий на основе текущей информации с целью достижения определенного, обычно оптимального, состояния системы при начальной неопределенности и изменяющихся условий работы». Применяя это определение к процессу обучения, можно сказать, что адаптация в обучающей системе – это процесс изменения параметров и структуры модели объекта (обучаемого) и обучающих воздействий на основе текущей информации, получаемой в ходе обучения, с целью достижения оптимального состояния объекта при его начальной неопределенности и изменяющейся среде. Начальная неопределенность связана с почти полным отсутствием в обучающей системе информации об обучаемом.

Обучающая система на рисунке 6 представлена как система управления сложным объектом – обучаемым с его моделью. Для более конкретной постановки задачи обучения необходимо формализовать:

- цель обучения Z^* ;
- обучающую информацию (ОИ), которую необходимо воспринять обучаемому и под воздействием которой у него должны сформироваться определенные знания и навыки, определяемые целью обучения;
- модель обучаемого;

- алгоритм обучения;
- критерий, позволяющий определить, достигнута ли поставленная цель.

Процесс обучения формализуется в виде последовательности сеансов (уроков), начинающихся в моменты времени t_0, t_1, \dots, t_n , в общем случае не равноотстоящие. В начальный момент времени объект (обучаемый) находится в некотором состоянии Y_0 . Требуется построить последовательность обучающих воздействий $\{U_n\}$, $n=0, 1, \dots$, которая переведет ученика в заранее заданное конечное состояние Y^* , причем процесс перевода должен быть, в определенном смысле, оптимальным. В задачах обучения лучшим следует считать тот алгоритм, который осуществляет процесс перевода за кратчайшее время.

Для определения эффективности необходимо ввести функцию качества Q обучения, которая должна зависеть от состояния объекта Y , и вычислять ее значения в дискретные моменты t_0, t_1, \dots, t_n :

$$Q_n = Q(Y_n), \quad (1.1)$$

где Y_n - состояние объекта в момент начала n -го сеанса обучения t_n .

Без ограничения общности можно считать, что

$$Q(Y^*) = Q^* = \min, \quad (1.2)$$

где уровень Q^* будет соответствовать абсолютной обученности.

Цель обучения Z^* , таким образом, состоит в минимизации функции качества Q с помощью U :

$$Q(Y) \rightarrow \min_{u \in U}. \quad (1.3)$$

Ввиду реальных свойств человеческой памяти уровень абсолютной обученности Q^* практически не достижим. Поэтому обучение следует заканчивать, когда критерий качества достигает некоторого заданного порога:

$$Q_n \leq \delta, \quad (1.4)$$

где $\delta > Q^*$ - величина, близкая к Q^* .

Цель обучения формализуется следующим образом:

$$Z^*: \begin{cases} Q(Y^*) \leq \delta; \\ T(Y^*) = \min; \end{cases} \quad (1.5)$$

где $T(Y^*)$ - время, за которое обучаемый достигает состояния Y^* .

При формализации ОИ необходимо рассматривать такие процессы обучения, в которых ОИ можно представить в виде конечного множества перенумерованных элементарных порций: $U = \{1, 2, \dots, N\}$. Из этого множества номеров на каждом n -ом сеансе с помощью алгоритма обучения строится подмножество:

$$U_n = \{u_1, u_2, \dots, u_{M_n}\}, u_i \neq u_j \text{ при } i \neq j, u_i \in U. \quad (1.6)$$

Это подмножество содержит M_n элементарных порций ОИ с соответствующими номерами.

В качестве модели обучаемого предлагается модификация адаптивной модели обучаемого Л.А. Растригина применительно к мультилингвистической адаптивно-обучающей технологии.

Состояние обучаемого на n -ом сеансе описывается вектором вероятностей незнания каждого из элементов ОИ:

$$Y_n = P_n = (p_1^n, p_2^n, \dots, p_N^n), \quad (1.7)$$

где p_i^n - вероятность незнания i -го элемента в n -й момент времени t_n .

Очевидно, что в момент заучивания порции ОИ на n -м сеансе ученик знает элементы данной порции с вероятностью единица, т.е. вероятности незнания элементов из U_n в момент t_n равны нулю. Однако с течением времени происходит их забывание. Вероятности незнания элементов ОИ изменяются по правилу

$$p_i^n = p_i(t_i^n) = 1 - e^{-\alpha_i^n t_i^n}, \quad (1.8)$$

где α_i^n - скорость забывания i -го элемента ОИ на n -м сеансе с учетом его связи с элементами ранее изучавшихся иностранных терминологий; t_i^n - время с момента последнего заучивания i -го элемента ОИ.

Скорость забывания каждого элемента уменьшается, если этот элемент выдается обучаемому для запоминания, и не изменяется, если он не заучивается :

$$\alpha_i^{n+1} = \begin{cases} \alpha_i^n, & \text{если } i \notin U_n, \\ \gamma' \alpha_i^n F_n^{NEG}, & \text{если } i \in U_n \text{ и } r_i^n = 1, \\ \gamma'' \alpha_i^n F_n^{NEG}, & \text{если } i \in U_n \text{ и } r_i^n = 0,5, \\ \gamma''' \alpha_i^n F_n^{NEG}, & \text{если } i \in U_n \text{ и } r_i^n = 0, n = 1, 2, \dots, \end{cases} \quad (1.9)$$

где α_i^1 – начальное значение скорости забывания, оцениваемое методом максимального правдоподобия, $0 < \alpha_i^1 < 1$, ($i = 1, 2, \dots, N$);

$\gamma', \gamma'', \gamma'''$ – параметры коррекции скоростей забывания, характеризующие индивидуальные особенности памяти обучаемого, оцениваемые методом максимального правдоподобия, $0 < \gamma' < 1$, $0 < \gamma'' < 1$, $0 < \gamma''' < 1$;

$R_n = (r_1^n, \dots, r_{Mn}^n)$ – множество ответов обучаемого на тесты:

$$r_i^n = \begin{cases} 1, & \text{если обучаемый не вспомнил термины на 2-х языках,} \\ 0.5, & \text{если не вспомнил термин хотя бы на одном языке,} \\ 0, & \text{если ни одного термина не забыл.} \end{cases} \quad (1.10)$$

Критерием качества обучения Q_n необходимо выбрать такой, который характеризует уровень обученности обучаемого. Для рассматриваемой задачи данный уровень характеризуется вероятностью незнания элемента ОИ, наугад выбранного из текста:

$$Q_n = \sum_{i=1}^N p_i(t_i^n) q_i \rightarrow \min; \quad (1.11)$$

где $p_i^n(t_i^n)$ – вероятность незнания i -го элемента ОИ;

q_i – относительная частота, выражающая долю лексической единицы в тексте, подвергшемуся статистической обработке при составлении частотного словаря, $0 < q_i < 1$,

$$q_i = \frac{q_i^{\max}}{V}; \quad (1.12)$$

где $q_i^{\max} = \max q\{q_{i1}, q_{i2}, \dots, q_{in}\}$ – абсолютная частота появления лексической единицы в тексте,

$q_{i1}, q_{i2}, \dots, q_{in}$ – частоты из мультилингвистического словаря.

Результатом решения данной задачи является локально-оптимальная порция ОИ U_n^* , которая выдается обучаемому на n -м сеансе обучения.

Для минимизации значения Q_n к концу сеанса обучения естественно в порцию U_n^* включать элементы ОИ, имеющие наибольшее значение произведения $p_i(t_i^n)q_i$, так как в результате их запоминания это произведение становится равным нулю и тем самым наибольшим образом снижает значение Q_n .

Таким образом, для обеспечения оптимального значения Q_n к концу n -го сеанса обучения, необходимо найти M_n максимальных членов суммы в критерии, индексы которых и определяют очередную порцию ОИ, выдаваемую ученику для запоминания. Процедура поиска индексов для терминов записывается следующим образом:

$$\begin{aligned} u_1 &= \arg \max_{1 \leq i \leq N} p_i(t_i^n)q_i; \\ u_2 &= \arg \max_{1 \leq i \leq N} p_i(t_i^n)q_i; \\ u_{M_n} &= \arg \max_{1 \leq i \leq N} p_i(t_i^n)q_i; \end{aligned} \tag{1.13}$$

$$i \neq u_j \quad (j = 1, 2, \dots, M_n - 1),$$

где $\arg \max \{a_i\} = i^*$ – индекс $i^* \in U$ максимального значения a_i , т.е. $a_{i^*} = \max a_i$, и, таким образом, $\{u_1, \dots, u_{M_n}\} = U^*$ – та порция ОИ, которая выдается для заучивания на n -м сеансе.

Объем этой порции зависит от T_n – продолжительности n -го сеанса обучения, или времени, отведенного на заучивание порции U_n . Предполагается,

что время заучивания i -го элемента прямо пропорционально вероятности его незнания, т.е. очевидно, что чем меньше вероятность незнания элемента, тем меньше времени необходимо на его заучивание. Тогда объем M_n очередной порции U_n определяется из следующего соотношения:

$$M_n = \max_{1 \leq M \leq N} \left\{ M : T_n \geq \kappa \sum_{i \in \{u_1, \dots, u_M\}} p_i(t_i^n) \right\}, \quad (1.14)$$

где κ - среднее время заучивания элемента ОИ при первом его предъявлении ученику;

u_1, \dots, u_M - номера элементов ОИ.

Параметр κ априорно неизвестен и поэтому должен оцениваться адаптивно в процессе обучения в зависимости от времени, затрачиваемого обучаемым на выполнение порции ОИ:

$$\kappa_{n+1} = \kappa_n + \mu (T'_n - T_n), \quad (1.15)$$

где μ - безразмерный коэффициент скорости адаптации,

T'_n - время, затраченное учеником на заучивание U_n .

Обучение заканчивается, когда Q_n достигает требуемого уровня обученности δ . Число сеансов обучения n , за которое достигается $Q_n \leq \delta$, определяет продолжительность обучения.

2 Оптимизация информационно-терминологического базиса мультилингвистической обучающей технологии

2.1 Информационно терминологический базис мультилингвистической обучающей технологии

При изучении любого иностранного языка всегда стоит задача запоминания больших объемов иностранной лексики. Под лексикой понимаются слова, словосочетания, а также грамматические формы, представляемые в виде словосочетаний. Проблема знания иностранной лексики наиболее остро встает перед специалистами, работающими с научной литературой на иностранном языке. При этом объем применяемых терминов ограничивается предметной областью, выбранной для изучения.

Для выделения из интересующих специалиста текстов необходимой лексики следует провести предварительную обработку этих текстов. А именно, следует составить частотный словарь слов и словосочетаний. Наличие такого словаря позволяет при обучении лексике учитывать статистические закономерности текстов. Со статистической точки зрения язык представляет собой большое количество редких событий (Закон Ципфа), в результате чего небольшое количество слов встречается очень часто, а подавляющее большинство слов имеют очень невысокую частоту. Естественно, что обучаемый должен быстрее и лучше усвоить ту лексику, которая в его текстах чаще всего встречается, т.е. следует учитывать частотные свойства текстов.

Целью применения информационной мультилингвистической адаптивно-обучающей технологии является интенсивное накопление специализированного иностранного словарного запаса студентов и специалистов, изучающих один или несколько иностранных языков для своих профессиональных целей. Основными компонентами средств ее поддержки являются электронные частотные словари, построенные по

мультилингвистическому принципу, и компьютерные системы, реализующие алгоритм обучения терминологической лексики.

Частотный словарь – это именно та информация, порции которой выдаются ученику. Обучаемый должен быстрее и лучше усвоить те слова и словосочетания, которые чаще всего встречаются в текстах узкой предметной области, т.е. необходимо учитывать частотные свойства текстов. После получения частотного словаря его необходимо разбить на учебные порции для наиболее эффективного усвоения материала.

2.2 Моделирование динамики изучения информационно-терминологического базиса с использованием цепей Маркова

В соответствии с принятым формализмом цепей Маркова, изучение информационно-терминологического базиса может рассматриваться как динамическая система, находящаяся в каждый из моментов tk - в одном из n состояний:

$$S_i(t_k) \in S(t_k) = \{S_1, \dots, S_n\}, t_k \in T.$$

переменная t_k , определяет номер шага в изучении базиса и не связана непосредственно со временем. Эта переменная принадлежит некоторому множеству рассматриваемых моментов времени T ,

Основное допущение, принятое в теории цепей Маркова - независимость вероятностей перехода из одного состояния в другое от предыстории процесса, т.е. состояния изменяются со временем случайным образом. Это изменение определяется матрицей переходных вероятностей

$$\begin{array}{c}
S_1 \quad S_2 \quad \dots \quad S_n \\
\begin{array}{c} S_1 \\ S_2 \\ \dots \\ S_n \end{array} \left[\begin{array}{cccc} p_{11}(t_k) & p_{12}(t_k) & \dots & p_{1n}(t_k) \\ p_{21}(t_k) & p_{22}(t_k) & \dots & p_{2n}(t_k) \\ \dots & \dots & \dots & \dots \\ p_{n1}(t_k) & p_{n2}(t_k) & \dots & p_{nn}(t_k) \end{array} \right]
\end{array} \quad (2.1)$$

Каждый элемент матрицы $p_{ij}(t_k)$ показывает вероятность того, что если система в момент t_k находилась в состоянии S_i , то в момент t_k+1 она окажется в состоянии S_j : $p_{ij}(t_k) = \Pr\{S(t_k+1) = S_j \mid S(t_k) = S_i\}$

Каждая строчка матрицы P соответствует состоянию, в котором процесс находится на данном шаге, а каждый столбец - состоянию, в которое переходит процесс в следующем шаге.

Процесс случайного перехода системы из одного состояния в другое называют цепью Маркова.

Переходы во все возможные состояния (в том числе в себя) образуют полную группу событий, поэтому $\sum_{j=1}^n p_{ij}(t_k) = 1$ для всех $i = 1, \dots, n, t_k \in T$.

В дальнейшем предполагается, что вероятности не зависят от времени, т.е. цепь Маркова однородна.

Пусть вектор-строка $X(t_k)=[x_1(t_k), \dots, x_n(t_k)]$ - описывает распределение вероятностей нахождения системы в соответствующих состояниях в момент t_k , то есть $x_i(t_k)$ - это вероятность того, что в момент t_k система находится в состоянии S_i . При этом $\sum_{i=1}^n x_i(t_k) = 1, t_k \in T$. Тогда по теореме об умножении вероятностей и с учетом основного свойства марковского процесса получим:

$$x_j(t_k+1) = \sum_{i=1}^n p_{ij}(t_k) x_i(t_k), \quad (2.2)$$

где $p_{ij}(t_k)$ выступают в роли условных вероятностей перехода в состояние S_j , при условии, что система находится в состоянии S_i .

В матрично-векторной форме (2.2) примет вид:

$$X(tk+1)=X(tk)P, \quad (2.3)$$

Должно быть также задано начальное условие: $X(t_0)=X_0$, которое определяет состояние процесса на начальном шаге изучения информационного базиса.

Вычисляя последовательно $X(t_1)$, $X(t_2)$, ..., $X(t_k)$ мы можем получить вероятностный прогноз графика построения системы базиса.

Множество состояний системы S подразделяется на множество невозвратных состояний S^1 и множество поглощающих состояний S^2 . Состояния, относящиеся к множеству S^2 , соответствуют завершению процесса. Поэтому, исключив из матрицы P строки и столбцы, соответствующие состояниям из S^2 и обозначив оставшуюся матрицу Q , можем вычислить так называемую фундаментальную матрицу цепи Маркова:

$$N = (I - Q)^{-1}, \quad (2.4)$$

где I -единичная матрица.

Каждый элемент n_{ij} матрицы N представляет собой среднее число пребываний процесса в состоянии S_i при старте из состояния S_l . В нашем случае старт всегда происходит из состояния S_l , поэтому достаточно рассматривать только первую строку матрицы N .

Зная n_{lj} , можно вычислить среднюю трудоемкость формирования структуры частотного словаря по формуле

$$\Theta_{\Sigma} = \sum_{j=1}^s n_{lj} \Theta_j \quad (2.5)$$

где - Θ_j трудоемкость j - го шага процесса изучения информационного базиса в часах.

В ряде случаев формирование словаря может интересовать оценка дисперсии трудоемкости курса. Для этой цели вычисляется матрица дисперсий числа пребывания процесса во множестве невозвратных состояний по формуле:

$$D = N(2N_{dg} - I) - N_{sg} \quad (2.6)$$

где индексы dg и sq обозначают соответственно выделение диагональных элементов матрицы N и возведение в квадрат каждого элемента этой матрицы.

В ряде случаев исследователя интересует не дисперсия, а среднеквадратичное отклонение числа пребывания процесса от среднего, которое вычисляется по известной формуле:

$$\sigma_{ij} = \sqrt{d_{ij}},$$

или в матричной форме

$$\sigma = \|\sigma_{ij}\|. \quad (2.7)$$

Рассмотрим модель динамики изучения информационно-терминологического базиса на основе цепей Маркова с дискретным временем. Представление процесса изучения базиса в виде цепи Маркова основана на ряде допущений. Кроме указанного выше основного допущения приняты следующие: допущение об однородности цепи Маркова (т.е. о независимости вероятностей от времени); о независимости трудоемкости шага изучения информационно-терминологического базиса от числа обращений к нему; о строгом следовании предписанному порядку выполнения шагов изучения информационно базиса.

В настоящей работе представлены следующие оценки:

- распределение вероятностей определения наборов блоков частотного словаря на каждом шаге процесса изучения информационно-терминологического базиса, в том числе вероятность завершения изучения базиса за заданное число шагов;

- распределение вероятностей различных вариантов завершения изучения информационно-терминологического базиса, если такая возможность предусмотрена;

- средняя трудоемкость построения базиса;

- дисперсия трудоемкости построения базиса.

Кроме того, возможны более тонкие исследования структуры мультилингвистического частотного словаря, например, зависимость общей трудоемкости изучения базиса от степени его дробления на блоки.

Исходная информация для модели включает список узлов (шагов процесса изучения), граф связи между ними, матрицу вероятностей перехода от узла к узлу и оценку средней трудоемкости каждого шага. Список шагов процесса изучения информационного базиса, трудоемкости их выполнения в часах и связи между ними определяются структурой мультилингвистического словаря.

Вероятности переходов могут первоначально оцениваться экспертно на основе опыта работы при формировании структуры мультилингвистического частотного словаря, а впоследствии уточняться по результатам оценки структуры базиса на основе собранных статистических данных. То же касается и оценок трудоемкости отдельных шагов изучения информационного базиса.

2.3 Оптимизация структуры информационно-терминологического базиса.

Уже отмечалась актуальность задачи оптимального дробления структуры информационно-терминологического базиса на модули. Таким образом, можно попытаться определить оптимальную степень дробления базиса на модули.

Для дальнейшего анализа его структуры, сделаем следующие дополнительные предположения:

1. Материал базиса достаточно однородный, что позволяет дробить его на модули произвольного объема.

2. Вероятность успешного выполнения контрольных мероприятий зависит от объема информационного модуля базиса, и при его увеличении она уменьшается.

Следуя часто используемой дидактической модели процесса изучения информационно-терминологического базиса, можно принять гипотезу об экспоненциальном характере зависимости вероятности успешного прохождения контрольных мероприятий от объема базиса. При этом, учитывая связи между словами, принимается гипотеза о зависимости той же вероятности от степени соответствия слов REL. Если объем информационного материала некоторого базиса составляет Θ часов, то вероятность успешного выполнения контрольных мероприятий может быть представлена в виде:

$$p(\Theta) = b \exp(-\lambda\Theta) + (1 - b)REL, \quad (2.8)$$

где λ (1/час) - константа, показывающая скорость снижения вероятности успешного завершения изучения информационного базиса в зависимости от его объема, а REL – степень соответствия слов между собой ($0 < REL < 1$).

Зависимость (2.8) соответствует интуитивному представлению об успешности процесса изучения информационно-терминологического базиса. Если объем информационного базиса мал ($\Theta \rightarrow 0$), то вероятность изучения базиса стремится к единице, а если он весьма велик ($\Theta \rightarrow \infty$), то вероятность его изучения стремится к нулю.

Величина λ зависит от вида информационно-терминологического базиса, она может быть оценена экспертно или экспериментально на основе опыта работы с определенным видом информационно-терминологических базисов. Остановимся на оценке этого параметра.

В простейшем случае достаточно одного эксперимента. Пусть процесс изучения информационного базиса определен объемом Θ часов. При этом оценка вероятности успешного его завершения составила p . Из (2.8) следует, что

$$\lambda = - \frac{\ln(p - (1 - b)REL)}{\Theta b} \quad (2.9)$$

Если число экспериментов больше одного, то эту задачу можно решить с использованием метода наименьших квадратов. Предположим, проведено m экспериментов по изучению базиса объемом $\Theta_1, \dots, \Theta_m$ часов и получены вероятности успешного завершения его изучения p_1, \dots, p_m . Тогда оценка параметра λ может быть получена по формуле:

$$\lambda = - \frac{\sum_{i=1}^m \ln(p_i - (1 - b)REL_i) \Theta_i b}{\sum_{i=1}^m \Theta_i^2} \quad (2.10)$$

Здесь не рассматриваются статистические оценки достоверности полученных результатов. Они могут быть получены на основе стандартных методик.

Рассмотрим цепь Маркова, описывающую вероятностный процесс изучения отдельного базиса (рис. 7).

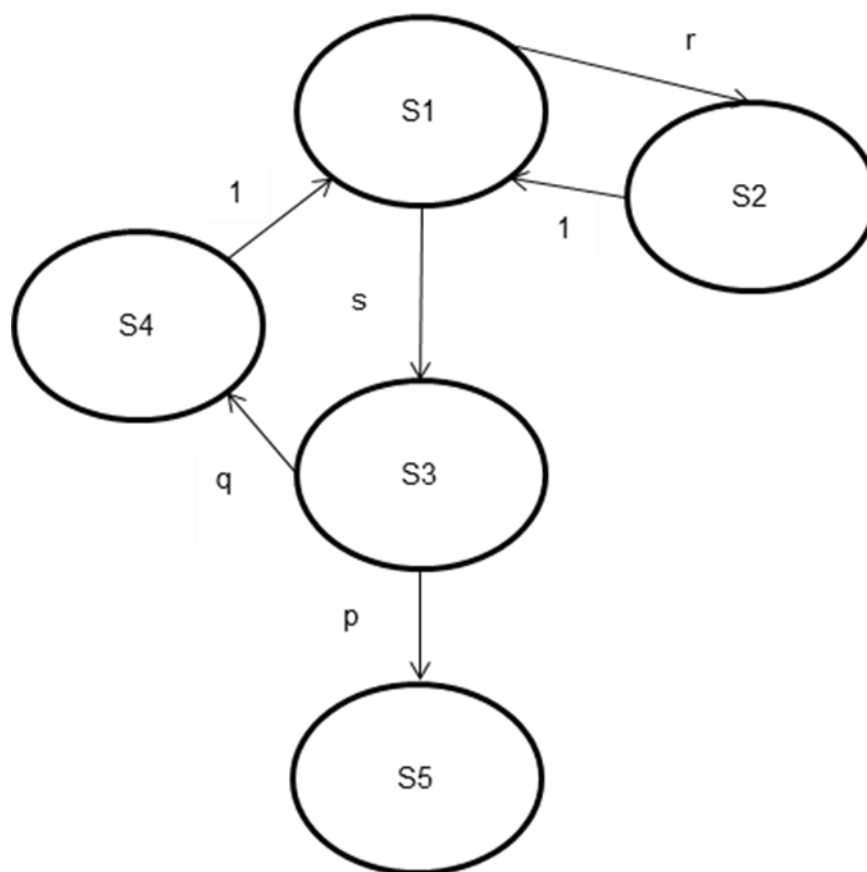


Рисунок 7 - Модель изучения базиса

В данной модели выделены следующие состояния:

S1 - изучение информационного материала;

S2 - получение консультаций при изучении слов;

S3 – выполнение контрольных мероприятий;

S4 – применение мнемотехник для работы над словами, в которых были допущены ошибки;

S5 - завершение изучения базиса.

Зададим вероятности переходов из одного состояния в другое:

r - вероятность обращения за справками или консультациями при изучении информационно-терминологического базиса;

s - вероятность перехода к контрольным мероприятиям;

p - вероятность успешного выполнения контрольных мероприятий и завершения изучения базиса;

q - вероятность неудачи при выполнении контрольных мероприятий и повторного изучения базиса.

В рассматриваемой схеме выполняются очевидные соотношения:

$$r+s=1, \quad p+q=1. \quad (2.11)$$

Матрица вероятностей переходов между состояниями:

$$P = \begin{array}{c} \begin{array}{ccccc} & s1 & s2 & s3 & s4 & s5 \\ \begin{array}{c} S1 \\ S2 \\ S3 \\ S4 \\ S5 \end{array} & \begin{bmatrix} 0 & r & s & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q & p \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \end{array} \quad (2.12)$$

Выделим в P матрицу вероятностей переходов между состояниями невозвратного множества, которая обозначается Q

$$Q = \begin{array}{c} \begin{array}{ccccc} & s1 & s2 & s3 & s4 \\ \begin{array}{c} s1 \\ s2 \\ s3 \\ s4 \end{array} & \begin{bmatrix} 0 & r & s & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & q \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \end{array} \quad (2.13)$$

Тогда фундаментальная матрица цепи Маркова определяется по формуле (2.4):

$$N = (I - Q)^{-1} = \begin{bmatrix} 1 & -r & -s & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -q \\ -1 & 0 & 0 & 1 \end{bmatrix}^{-1}, \quad (2.14)$$

где I - единичная матрица размерности 4x4.

Поскольку процесс всегда стартует из состояния Si , нам достаточно определить только первую строчку матрицы N ; ее элементы обозначим $n1, \dots, n4$. Проведем вычисления обратной матрицы $(I - Q)^{-1}$, получим оценки среднего числа попыток изучения всех компонентов информационного базиса - среднее число пребываний процесса соответственно в состояниях $\{S1, \dots, S6\}$ при старте из состояния $S1$:

$$n_1 = \frac{1}{ps}; \quad n_2 = \frac{r}{ps}; \quad n_3 = \frac{1}{p}; \quad n_4 = \frac{q}{p}; \quad (2.15)$$

Здесь:

$n1$ - среднее число попыток изучения информационного материала;

$n2$ - среднее число обращений за справками и консультациями при изучении информационного материала;

$n3$ - среднее число попыток выполнения контрольных мероприятий;

$n4$ - среднее число применений мнемотехник для работы над словами, в которых были допущены ошибки;

Обозначим трудоемкости изучения всех компонентов информационного базиса (в часах):

Θ - трудоемкость изучения информационного материала;

$\Theta1$ - трудоемкость получения справок и консультаций при изучении информационного материала (состояние $S2$);

$\Theta2$ - трудоемкость выполнения контрольных мероприятий (состояние $S3$);

$\Theta3$ - трудоемкость выполнения мнемотехник (состояние $S4$).

Суммарная трудоемкость прохождения информационного базиса с учетом (2.15) определяется выражением:

$$R = n_1 \cdot \Theta + n_2 \cdot \Theta_1 + n_3 \cdot \Theta_2 + n_4 \cdot \Theta_3 = \frac{1}{p} \left[\frac{1}{s} (\Theta + r\Theta_1) + \Theta_2 + q\Theta_3 \right] \quad (2.16)$$

Рассмотрим оценки параметров, входящих в формулу (2.16).

Вероятность r определяются экспертно или на основе обработки результатов изучения информационно-терминологического базиса. Вероятность p определяется формулой (2.11).

Трудоемкость Θ определяется объемом информационного базиса, трудоемкость $\Theta 1$ может быть взята как некоторая доля величины Θ :

$$\Theta 1 = k1 \Theta, \quad (2.17)$$

где $0 < k1 < 1$ – доля затрат на получение справок и консультаций, определяемая экспертно или на основе статистики.

Трудоемкость $\Theta 2$ может быть представлена как сумма постоянной и переменной составляющих:

$$\Theta 2 = m + k2 \Theta, \quad (2.18)$$

где m (часов) – трудоемкость работы по выполнению контрольных мероприятий и не зависящая от размеров базиса, а Rel – это экспертная оценка снижения трудоемкости за счет соответствия слов.

$0 < k2 < 1$ - доля затрат на выполнение контрольных мероприятий.

Трудоемкость $\Theta 3$ можно представить, как сумма доли трудоемкости Θ и постоянная составляющая:

$$\Theta 3 = w + k3 \Theta, \quad (2.19)$$

где w (часов) – трудоемкость работы по изучению мнемотехник и не зависящая от размеров базиса.

$0 < k3 < 1$ - доля затрат на применение мнемотехник.

В результате подстановки значений параметров (2.17), (2.18) формула (2.16) приобретет вид:

$$R = \frac{1}{be^{-\lambda \Theta} + (1-b)REL} \left[\left(\frac{1}{s} (1 + k1r) + k2 + qk3 \right) \Theta + m + qw \right] = \quad (2.20)$$

$$R1 + R2 + R3$$

причем

$$R_1 = \frac{1}{be^{-\lambda \Theta} + (1-b)REL} [1 + k_1 r] \frac{\Theta}{s} - \text{затраты на изучение базиса.}$$

$$R_2 = \frac{1}{be^{-\lambda \Theta} + (1-b)REL} (k_2 \Theta + m) - \text{затраты времени на выполнение}$$

контрольных мероприятий.

$$R_3 = \frac{1}{be^{-\lambda \Theta} + (1-b)REL} [qk_3 \Theta + qw] - \text{затраты времени на выполнение}$$

мнемотехник.

Обозначим

$$\frac{1}{s}(1 + k_1 r) + k_2 + qk_3 = k$$

С учетом (2.20) выражение для суммарной трудоемкости (2.19) упростится

$$R = \frac{1}{be^{-\lambda \Theta} + (1-b)REL} (k\Theta + m + qw) \quad (2.21)$$

2.4 Задача об оптимальной разбивке информационно-терминологического базиса на модули

Рассмотрим теперь задачу разбивки некоторого информационного базиса с общим объемом материала Θ часов на n модулей, каждый из которых имеет объем Θ_i часов, так что

$$\Theta = \sum_{i=1}^n \Theta_i. \quad (2.22)$$

Трудоемкость изучения каждого модуля в соответствии с (2.21) составит:

$$R_i = \frac{1}{be^{-\lambda_i \Theta_i} + (1-b)REL} (k_i \Theta_i + m_i + qw_i) \quad (2.23)$$

а общая трудоемкость изучения базиса

$$R = \sum_{i=1}^n R_i = \sum_{i=1}^n \frac{1}{be^{-\lambda_i \Theta_i} + (1-b)REL} (k_i \Theta_i + m_i + q w_i) \quad (2.24)$$

Рассмотрим предельные случаи. При $n = 1$ получаем формулу (2.21). При $n \rightarrow \infty$ $\Theta_i \rightarrow 0$, $\exp(\lambda_i \Theta_i) \rightarrow 1$, и из (2.24) следует, что $R \rightarrow \infty$. Таким образом, функция R имеет, по крайней мере, один минимум.

Выбирая количество модулей n и их объем Θ_i , можно добиться наименьшей общей трудоемкости изучения базиса.

Математически эта задача формулируется следующим образом: задан критерий (2.24) при условиях (2.22). Требуется найти такие n и $\{\Theta_1, \dots, \Theta_n\}$, чтобы обеспечить оптимальные значения критерия (2.24):

$$\hat{R} = \min_{n, \Theta_1, \dots, \Theta_n} R \quad (2.25)$$

Задача (2.25) представляет собой задачу оптимизации нелинейного критерия при ограничениях на переменные. В аналитическом виде эта задача решается достаточно сложно, а численно может быть решена путем перебора вариантов разбивки базиса на модули, если дополнительно задать процедуру формирования таких вариантов.

2.5 Формирование модулей по семантическому принципу

В данной работе связь между словами определяется по семантическому принципу. Основываясь на этом принципе, был разработан алгоритм формирования модуля, с помощью которого можно количественно оценить силу связи между словоформами в рамках исследуемого текста. Также стоит отметить, что в данный алгоритм встроен алгоритм определения семантической близости ключевых слов по их окружению в тексте (алгоритм Гинзбурга).

Подробно алгоритм формирования модуля представлен ниже:

1. Переменная E равна 1, переменная $LangCount$ равна 4.

2. Находим в тексте Т для каждой словоформы а ее относительную частоту – $OTЧ_e(a)$ для $E_{\text{того}}$ языка. В результате получаем частотный словарь для текста Т на $E_{\text{том}}$ языке.
3. $E = E + 1$
4. Если E больше LangCount, то переходим к шагу 5. Иначе, к шагу 2.
5. Исключаем все слова из базиса, для которых не определены все относительные частоты для каждого из языков.
6. Количество слов в базисе, который хранится в базе данных, записывается в переменную SizeOfBasis, переменная ModNum равна 0.
7. Для каждого слова в базисе определяем его среднюю ОТЧ:

$$OTЧ(a) = \frac{\sum_{i=1}^{LangCount} b_i * OTЧ_i(a)}{LangCount},$$

где $\sum_{i=1}^{LangCount} b_i = 1$; b_i – коэффициент важности языка

8. Сортируем массив слов в порядке убывания по характеристике ОТЧ (а)
9. Переменная I равна 1.
10. Если I больше SizeOfBasis, то переходим к шагу 25. В противном случае берется слово из базы данных под индексом I. К значению ModNum прибавляем 1.
11. Если значение поля mod в базе данных для слов под индексом I равно 0, то устанавливаем это значение равным ModNum, слова записывается в массив MainWords размера LangCount, иначе к значению I прибавляется 1 и переходим к шагу 11.
12. Переменная E=1
13. Находим в Т те части этого текста одного и того же размера, которые содержат заданное слово I для Eтого языка – предложения с этим словом. Для совокупности всех этих предложений Т* построим частотный словарь V(Т*), содержащий абсолютную и относительную частоты. Относительную частоту словоформы а в V(Т*) обозначим $OTЧ_i^*(a)$.
14. $E = E + 1$
15. Если E больше LangCount, то переходим к шагу 16. Иначе к шагу 13.

16. Переменная $J = I + 1$
17. Если J больше $SizeOfBasis$, то переходим к шагу 10. Если нет, то из базы данных берутся слова из строки под индексом J . Переменная $SumLangKof$ равна 0.
18. Переменная $S = 1$, $OTCH^* = 0$
19. Если S больше $LangCount$, то переходим к шагу 23. Иначе к шагу 20.
20. Если $OTCH_s^*(I)$ существует, то:

$$OTCH^* = OTCH_s^*(I) * b_s$$

В этом случае переходим к шагу 21, иначе к 24.

21. $S = S + 1$, переходим к шагу 19
22. Сравниваем полученные относительные частоты в T и T^* : Вводим Индекс значимости словоформы a в контексте слова I ($ИнЗ(a)$), вычисляемый по формуле:

$$ИнЗ(a) = \frac{OTCH^*(a)}{OTCH(a)}$$

если $ИнЗ(a) >$ порогового значения, переходим к шагу 23, иначе - 24

23. Значению поля $mod\ j_{того}$ слова присваиваем $modNum$.

24. $J = J + 1$

25. Выход

Далее после того, как модули сформированы, рекомендуется применить алгоритм формирования блоков из заранее заданных модулей.

2.6 Алгоритм случайного поиска

1. Аналитиком задается количество групп при объединении NR и количество вариантов объединений слов в группы NG , объем курса Si , вероятность обращения за справками r , затраты при обращении за справками $k1$, затраты при выполнении контрольных работ $k2$, затраты на применение мнемотехник $k3$, сложность усвоения материала l , трудоемкость проведения контрольных работ m , вероятность успешного завершения изучения базиса p ,

организационные издержки при изучении с использованием мнемотехник w , коэффициент важности объема базиса b .

2. Переменная s равна $1-r$, переменная q равна $1-p$.

3. Переменная k равна:

$$k = \frac{1}{s}(1 + k_1 r) + k_2 + qk_3$$

4. Переменная WordsCount равна количеству слов в базисе.

5. Создаем массив effc размера WordsCount-1.

6. Переменная H равна 2.

7. Создаем массив Group размера H .

8. Если H больше WordsCount, то переходим к шагу 27.

9. Переменная REff равна 0.

10. Переменная J равна 1.

11. Если J больше NR , то переходим к шагу 25. Создаем массив WordsSi размера H .

12. Переменная SiSum равна 0. Переменная RSum равна 0.

13. Переменная F равна 1.

14. Если F больше H , то переходим к шагу 22.

15. Генерируем случайным образом значение, присваиваемое переменной I .

16. Проверяем, совпадает ли I со значениями элементов массива Group. Если нет, то записываем I в массив. Иначе переходим к шагу 15.

17. Переменной VOL присваиваем количество слов, входящих в блок I .

18. В массив ModSi[F] записываем значение, равное VOL, деленной на WordsCount и умноженной на Si.

19. К RSum добавляем значение:

$$RSum = \frac{1}{be^{-I \text{ModSi}[F]} + (1-b)REL} (k \text{ModSi}[F] + m + qw)$$

20. К SiSum добавляем значение WordsSi[F].

21. К F прибавляем 1 и переходим к шагу 14.

22. Переменной RAll присваиваем:

$$RAll = \frac{1}{be^{-lSiSum} + (1-b)REL} (kSiSum + m + qw)$$

23. К REff добавляем значение RAll/RSum.

24. К J добавляем 1 и переходим к шагу 11.

25. В effc[H-2] записываем REff/NR.

26. Увеличиваем H на 1 и возвращаемся к шагу 8.

27. Ищем минимальное значение effc, индекс соответствующего элемента записываем в переменную min.

28. Переменная BlocksCount равна количеству блоков.

29. Создаем массив AllEffic размера NG.

30. Создаем массив Groups размером в NG строк и BlocksCount столбцов.

31. Переменная L равна 1.

32. Если L больше NG, то переходим к шагу 39.

33. Переменная S равна 1. Rsum равна 0.

34. Если S больше BlocksCount, то переходим к шагу 32.

35. Генерируем случайным образом значение, присваиваемое переменной I.

36. Проверяем, совпадает ли I со значениями элементов массива Groups в строке NG. Если нет, то записываем I в массив. Иначе переходим к шагу 35.

37. Прибавляем 1 к S и переходим к шагу 34.

38. Прибавляем 1 к L и переходим к шагу 32.

39. Переменная L равна 1.

40. Если L больше NG, то переходим к шагу 47.

41. Переменная K равна 1.

42. Если K больше BlocksCount, то переходим к шагу 40.

43. Считаем RAll, аналогично предыдущему, для групп по min элементов в строке и прибавляем значение к Rsum.

44. Rsum записываем в AllEffic[L].

45. Прибавляем 1 к K и переходим к шагу 42.

46. Прибавляем 1 к L и переходим к шагу 40.

47. Находим минимальный элемент AllEffic и номер элемента записываем в переменную N. Строка под номером N в массиве AllEffic является оптимальным объединением в модуль.

48. Выход.

3 Система мультилингвистического обучения

3.1 Рекомендуемые требования к аппаратному обеспечению

Тактовая частота процессора 300 MHz или более, от 128 MB RAM для Windows 2000/XP или Linux, от 512 MB RAM для Windows Vista/Seven;

3.2 Требования к программному обеспечению

- Операционная система Microsoft Windows 2000/XP/Vista/Seven x32|x64 или Linux+Wine);
- 12 MB свободной дисковой памяти на системном диске;
- Программа Microsoft Access 2010;
- Утилита управления СУБД - IB Expert 2+ (Windows), либо Flamerobin (Linux).

3.3 Установка программы

В процессе работы программа сохраняет информацию на носитель, поэтому некоторые функции программы не будут работать, если она будет находиться на CD-дисках. Для установки программы скопируйте папку с программой на жесткий диск. В комплект программы входят:

исполняемый файл MultiSystem.exe
словарь Example.accdb

3.4 Работа мультилингвистической обучающей системы

При запуске программы появляется следующее окно(рис. 8).

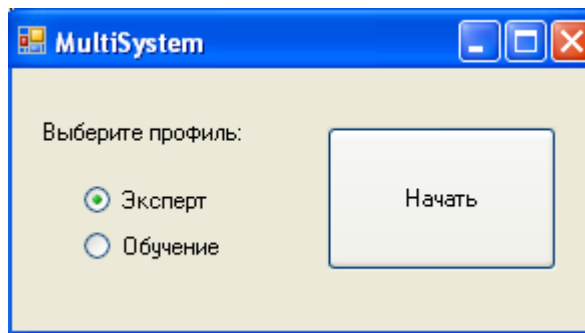


Рисунок 8 – Окно выбора режима работы

Программа имеет два режима работы:

- режим «Эксперт»;
- режим «Обучение».

Режим «Эксперт» включает в себя настройку параметров обучения, работу с базами данных, выбор типа контрольных мероприятий, анализ оптимального размера модуля, разбиение на модули.

В режим «Обучение» входят изучение модуля и контрольные мероприятия.

3.5 Режим «Эксперт»

При выборе режима «Эксперт», пользователь вводит текст, на основе которого будет строиться обучение (рис 9).

Form1

Текст:

disease, a thousand times more deadly than any they knew, arising and killing by the hundreds of millions and even by the billion. You see, the micro-organic world remained a mystery to the end. They knew there was such a world, and that from time to time armies of new gems emerged from it to kill men.

"And that was all they knew about it. For all they knew, in that invisible micro-organic world there might be as many different kinds of gems as there are grains of sand on this beach. And also, in that same invisible world it might well be that new kinds of gems came to be. It might be there that life originated—the 'abysmal fecundity,' Soldervetzsky called it, applying the words of other men who had written before him...."

It was at this point that Hare-Lip rose to his feet, an expression of huge contempt on his face.

[Illustration: Granser, you make me sick with your gabble 071]

"Granser," he announced, "you make me sick with your gabble. Why don't you tell about the Red Death? If you ain't going to, say so, an' we'll start back for camp."

The old man looked at him and silently began to cry. The weak tears of age rolled down his cheeks and all the feebleness of his eighty-seven years showed in his grief-stricken countenance.

"Sit down," Edwin counselled soothingly. "Granser's all right. He's just gettin' to the Scarlet Death, ain't you, Granser? He's just goin' to tell us about it right now. Sit down, Hare-Lip. Go ahead, Granser."

Обработать текст

Выход

Результат:

№	Слово	ОТЧ
1	the	0,0638297...
2	and	0,0381851...
3	of	0,0308581...
4	a	0,0277582...
5	was	0,0198675...
6	to	0,0164858...
7	in	0,0152176...
8	it	0,0131041...
9	that	0,0114132...
10	he	0,0111314...
11	his	0,0105678...
12	i	0,0101451...

Отобрать ТОП : 20

Отобрать

Рисунок 9 – Работа с текстом. Определение частотных характеристик всех словоформ.

Кнопка «Обработать текст» вычисляет частотные характеристики для каждого слова. Отобрать ТОП – выбор нужного количества базовых слов с наибольшей частотой, на основе которых будут строиться модули (рис. 10).

Form2

Список отобранных слов:

Назад

Слово	ОТЧ
my	0,001972664506...
world	0,001972664506...
see	0,001972664506...
which	0,001972664506...
great	0,001972664506...
death	0,001972664506...
people	0,001972664506...
called	0,001831759898...
good	0,001831759898...
time	0,001831759898...
fingers	0,001831759898...
these	0,001831759898...

Расчеты:

Расчеты для слова people

Новый текст:

Four million people lived in San Francisco then. "Where four million people disported themselves, the wild wolves roam to-day, and the savage progeny of our loins, with prehistoric weapons, defend themselves against the fanged despoilers.

"There were very many people in the world in those days.

"There were four million people in San Francisco—four teeth. Yes, my boy, all those people lived right here in San Francisco. And at one time or another all those people came out on this very beach—more people than there are grains of sand. And across the bay—where we camped last year, even more people lived, clear from Point Richmond, on the level ground and on the hills, all the way around to San Leandro—one great city of seven million people.

"The world was full of people. And the more food there was, the more people there were. Eight crab-shells there, ves, eight billion people were alive on the earth when the

Слова:

Слово	ОТЧ	ОТЧ*	ИНЗ
a	0,027758207693...	0,008771929824...	0,316012111497...
i	0,010145131745...	0,004385964912...	0,432322124756...
you	0,009017894885...	0,004385964912...	0,486362390350...
as	0,007467944201...	0,004385964912...	0,587305527970...
was	0,019867549668...	0,013157894736...	0,662280701754...
that	0,011413273214...	0,008771929824...	0,768572666233...
for	0,005636184303...	0,004385964912...	0,778179824561...
with	0,005636184303...	0,004385964912...	0,778179824561...
to	0,016485839086...	0,013157894736...	0,798133153396...
from	0,005495279695...	0,004385964912...	0,798133153396...
had	0,005213470480...	0,004385964912...	0,841275486012...
of	0,030858109060...	0,026315789473...	0,852799807738...

Рисунок 10 – Отбор языкового материала. Поиск контекста.

После того, как материал отобран, пользователь подключает базу данных с переводом на основных слов на другие языки. Редактирование базы данных осуществляется в окне «База данных» (рис. 11).

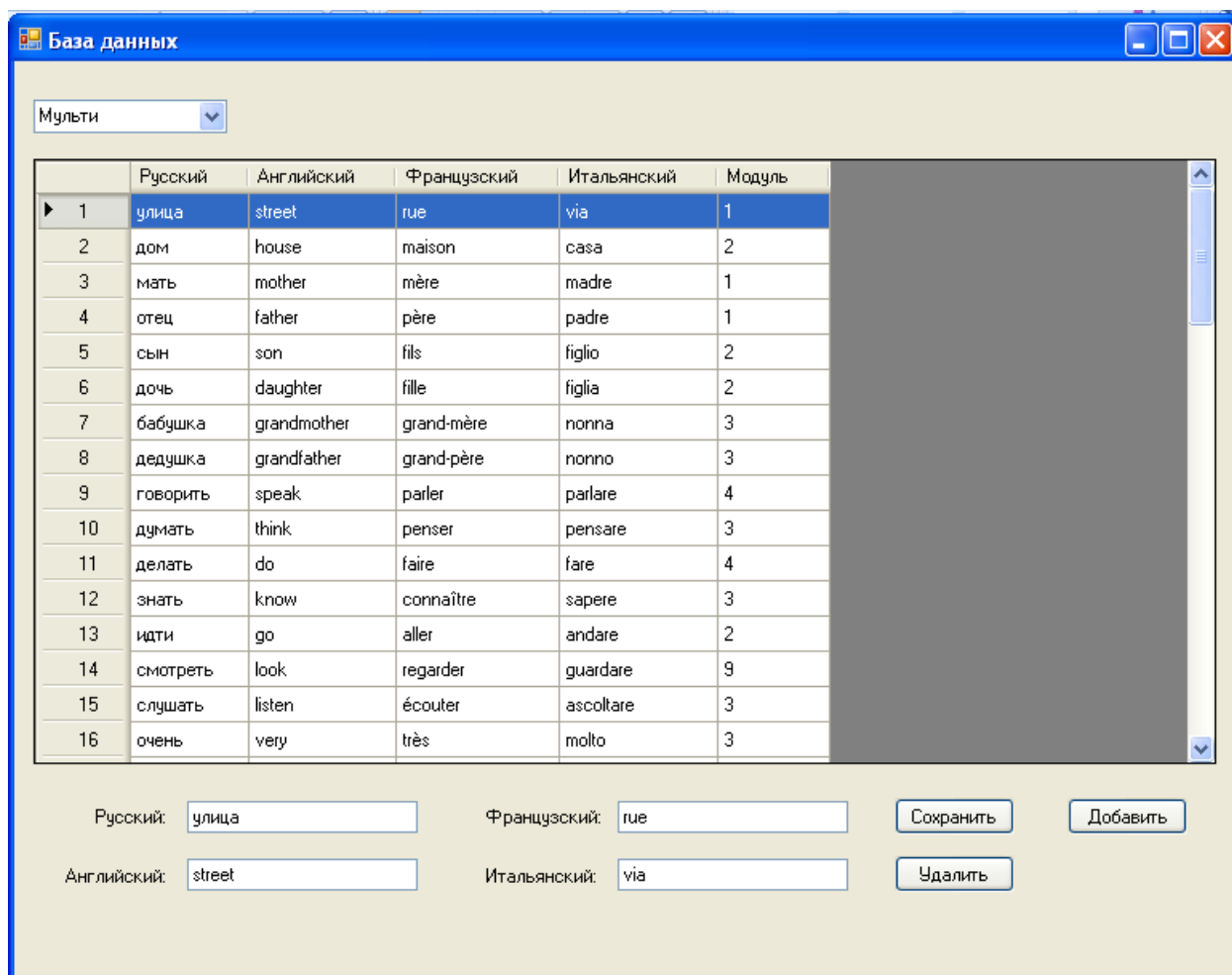


Рисунок 11 – Окно базы данных

Пользователь может добавлять новые слова, нажав на кнопку «Добавить». Откроется окно добавления слова (рис. 12).

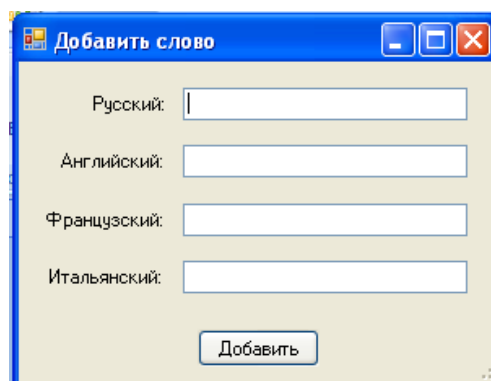


Рисунок 12 – Окно добавления слова

Пользователь также может редактировать уже имеющиеся слова в базе при выбранном пункте «Мульти» выпадающего списка. Добавление новых обучаемых производится в пункте «Логин».

Далее на вкладке «Настройки» эксперту необходимо задать максимальное число ошибок, которое разрешено допустить при прохождении контрольных мероприятий. Также требуется выбрать один из трех типов контрольных мероприятий:

- тест;
- тест с переменной языка;
- заполнение.

Подробнее каждый из типов контроля будет рассмотрен в режиме «Обучение».

Кроме того, для дальнейших расчетов при разбиении базиса необходимо заполнить параметры обучения (определяются экспертами):

- объем курса (ч.);
- вероятность обращения за справками;
- затраты при обращении за справками;
- затраты при выполнении контрольных работ;
- затраты на применение мнемотехник;
- сложность усвоения материала (1/час);
- трудоемкость проведения контрольных работ (ч.);
- вероятность успешного завершения изучения базиса;
- коэффициент важности объема базиса.

Закладка «Разбиение» (рис. 13) открывает страницу, на которой нужно задать дополнительные параметры:

- доля важности каждого языка при разбиении;
- коэффициент схожести слов;
- количество случайных разбиений.

После задания всех параметров необходимо нажать кнопку «Разбить», после чего словарь будет разбит на модули. Также станет активен выпадающий список «Модуль», с помощью которого можно показать в таблице только модуль под выбранным номером.

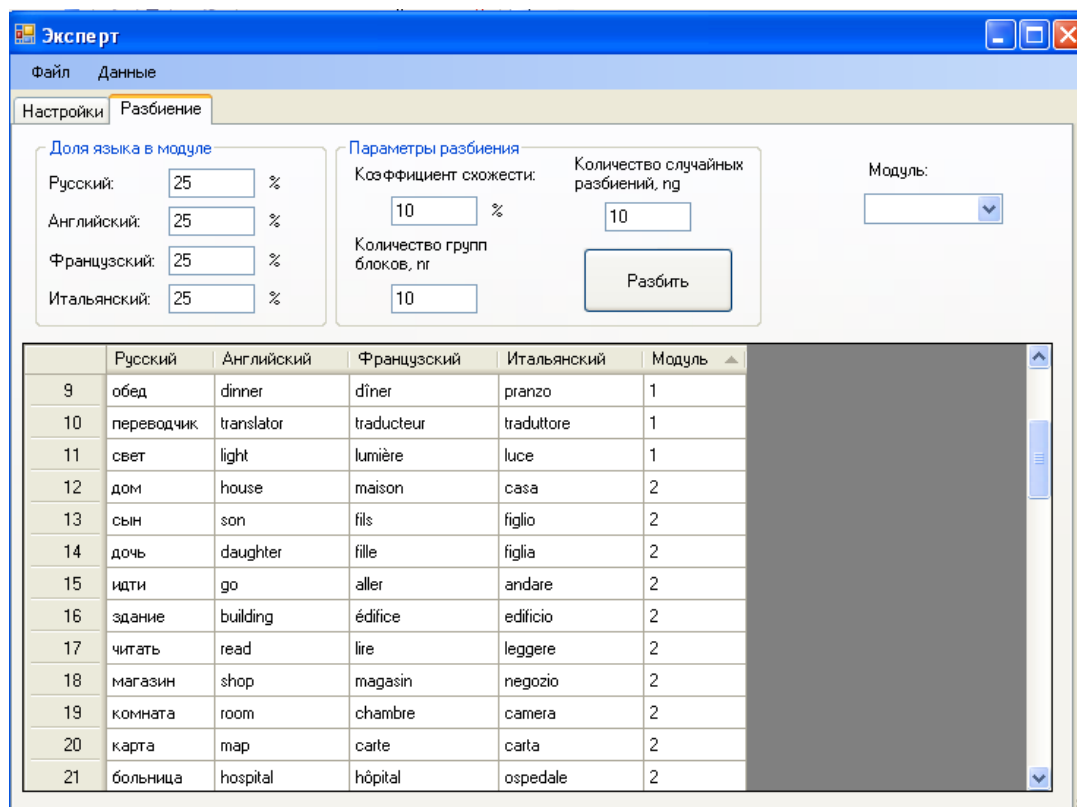


Рисунок 13 – Окно разбиения словаря на модули

3.6 Режим «Обучение»

При выборе режима «Обучение», пользователь переходит в окно изучения модулей (рис. 14).

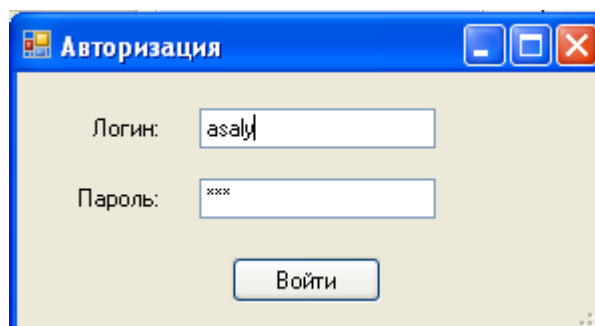


Рисунок 14 – Окно авторизации пользователя

Для продолжения изучения модулей и для отслеживания прогресса пользователю необходимо пройти аутентификацию. После этого можно приступать к изучению модуля.

По нажатию на кнопку «Начать обучение», обучаемому выводится модуль на котором он остановился (рис. 15).

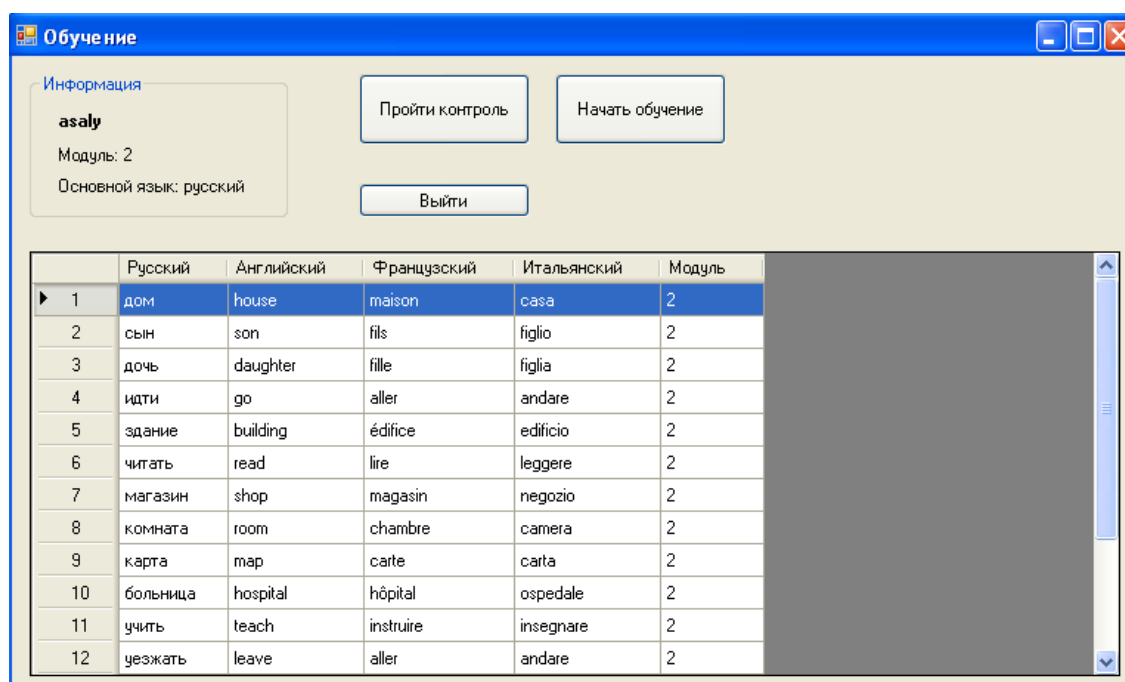


Рисунок 15 – Окно изучения модуля

После того как все слова изучены, обучаемый, для перехода на следующий модуль должен пройти контрольное мероприятие одного из трех видов выбранного экспертом. Для этого необходимо нажать кнопку «Пройти контроль». После этого появится окно с выбранным экспертом контрольным

мероприятием, а также информационные строки текущего прогресса и допущенных ошибок. Возвращения к обучению текущего модуля возможно лишь в случае превышения лимита ошибок в контрольном мероприятии, или же закрытием окна «обучение».

Типы контрольных мероприятий:

1) Тест – самый простой из типов и наименее адекватный, так как не происходит проверки знания каждого языка в отдельности. Обучаемому выводится слово на основном и предоставляется на выбор четыре варианта ответов, содержащие по три перевода на разные языки (рис. 16).

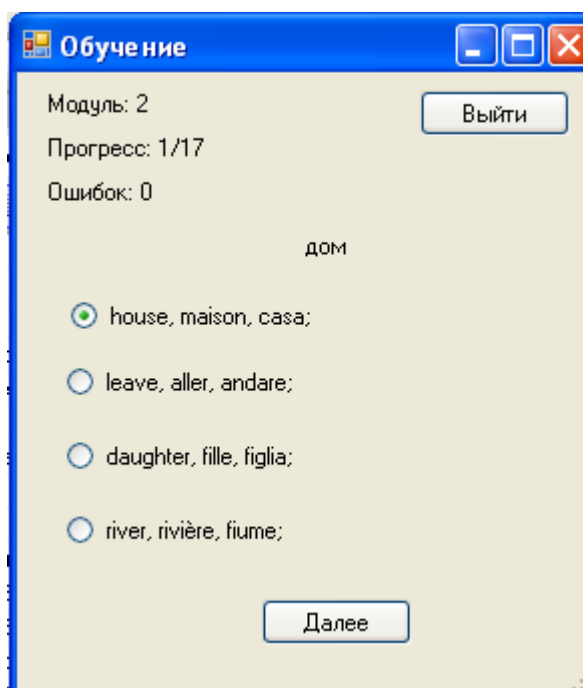


Рисунок 16 – Окно контрольных мероприятий типа тест

2) Тест с переменной языка – является наиболее оптимальным по сложности и адекватности проверки знаний среди трех представленных типов контрольных мероприятий, так как происходит проверка знания каждого языка в отдельности. Обучаемому выводится слово на основном языке (в скобках указано на какой язык требуется перевести) и предоставляется на выбор четыре варианта перевода (рис. 17).

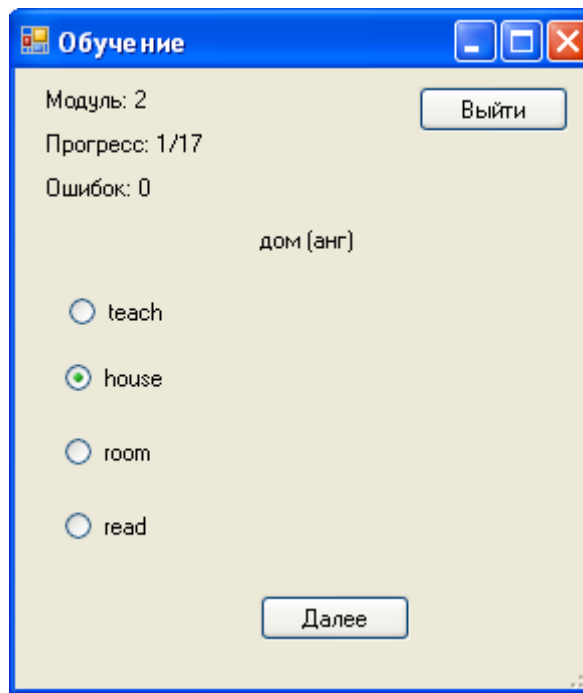


Рисунок 17 - Окно контрольных мероприятий типа тест с переменной языка

3) Заполнение – наиболее сложное контрольное мероприятие, которое является наиболее адекватным из представленных. Происходит проверка знания перевода каждого слова на каждый язык. При этом, проверяется не только визуальное восприятие слова, но и проверка знания орфографии. Обучаемому выводится слово на основном языке, после чего необходимо ввести перевод на каждом из изучаемых языков (рис. 18).

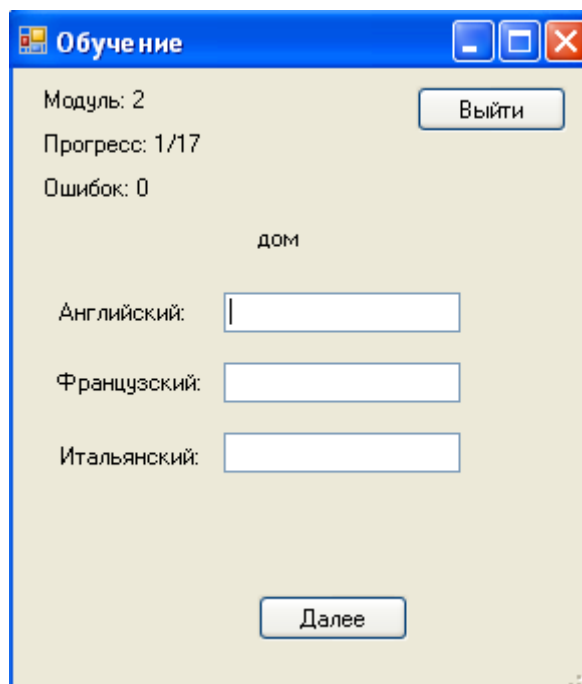


Рисунок 18 - Окно контрольных мероприятий типа заполнение

По окончании прохождения контрольного мероприятия происходит проверка количества ошибок, после чего обучаемый переходит к изучению следующего модуля при успешном прохождении контрольного мероприятия (рис. 19), или же возвращается к изучению текущего модуля при превышении допустимого числа ошибок.

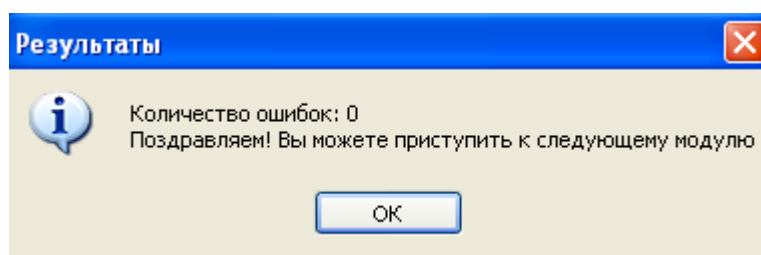


Рисунок 19 – Окно результатов контрольного мероприятия

ЗАКЛЮЧЕНИЕ

В ходе разработки системы мультилингвистического обучения с использованием оптимальной разбивки базиса на модули были достигнуты следующие основные результаты:

- проведено изучение мультилингвистической обучающей технологии;
- сформулированы задачи определения оптимальной структуры базиса с учетом схожести слов в каждом языке, и важности каждого языка в модуле;
- разработана система, позволяющая детально исследовать структуру базиса и оптимизировать разбивку словаря на модули;
- разработан модуль обучения на основе оптимальной разбивки словаря и выбранных методов контроля.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Савельев, А.Я. Подготовка информации для автоматизированных обучающих систем / А.Я. Савельев, В.А. Новиков, Ю.И. Лобанов. – Москва : Высшая школа, 1996. – 123 с.
- 2 Семенов, В.В. Информационные основы кибернетической компьютерной технологии обучения // Информатика и вычислительная техника. – 1997. – №3. – С. 37-40.
- 3 Мельников, А.В. Принципы построения обучающих систем и их классификация. / А.В. Мельников, П.Л. Цытович // Педагогические и информационные технологии в образовании. – 2001. – №4. – С. 26-28.
- 4 Голубева, Т.И. Применение информационных технологий в обучении иностранному языку : учебное пособие / Т.И. Голубева, Репина, С.О. – Оренбург : ГОУ ОГУ, 2004. – 9-13 с.
- 5 Растрин, Л.А. Адаптивное обучение с моделью обучаемого: учебник для вузов / Л.А. Растрин. – Рига : Зинатне, 1988. – 19-32 с.
- 6 Соловов, А. Дискретные математические модели в исследовании процессов автоматизированного обучения [Электронный ресурс] / А. Соловов, 2001. – 11 с. – Режим доступа: http://ifets.ieee.org/russian/depositary/v4_i2/html/3.html.
- 7 Растрин, Л.А. Адаптация сложных систем: учебник для вузов / Л.А. Растрин. – Рига : Зинатне, 1981. – 11-22 с.
- 8 Кривошеев, А.О. Разработка и использование компьютерных обучающих программ // Информационные технологии, – 1996. – №2. – С. 14-20.
- 9 Лесков, В.О. Два блока частотного словаря: значение и организация // Успехи современного естествознания. – 2009. – №4. – С.29-32.
- 10 Ковалев, И.В. Подготовка информации для автоматизированных обучающих систем : учебное пособие / И.В. Ковалев, А.А. Ступина, Е.А. Суздалева. – Томск, 2001. – 98-99 с.

- 11 Кларлащук, В.И. Обучающие программы. – Москва : Солон-Р, 2001. – 12 с.
- 12 Карасева М.В. Компьютерная система изучения терминологической лексики: мультилингвистический подход / М.В. Карасева, Т.А. Ковалев, Е.А. Суздалева // Информационные технологии в инновационных проектах – Ижевск: Изд-во Механического завода, 2000. – С.39-40.
- 13 Гутгарц, Р.Д. Компьютерная технология обучения // Информатика и образование. – 2009. – №5. – С. 24-25.
- 14 Карасева М.В. Автоматизация формирования информационной базы мультилингвистической адаптивно-обучающей технологии / М.В. Карасева. – Красноярск : СибГАУ, 2007. – Вып. 4(17). – С. 117-124.
- 15 Брусиловский П.Л. Интеллектуальные обучающие системы. // Информатика. Информационные технологии. Средства и системы. – 1990. – №2. – С. 22.
- 16 Бовтенко, М.А. Компьютерная лингводидактика : учебное пособие / М.А. Бовтенко. – Новосибирск : Изд-во НГТУ, 2000. – 124 с.
- 17 Крёнке Д. Теория и практика построения баз данных 9-е изд. – СПб. : Питер, 2005. – 859 с.
- 18 Ковалев И.В. Мультилингвистический метод изучения иностранной терминологической лексики на базе мнемотехнического подхода. – Тамбов : Издательство ТГУ им. Г.Р. Державина, 2001. – С. 43-44.
- 19 Калянов Г.Н. Сравнительный анализ структурных методологий // Системы управления базами данных. – 1997. – №5. – С. 75-77.
- 20 Воронина, И.Е. Алгоритмы определения семантической близости ключевых слов по их окружению в тексте / И.Е. Воронина // Вестник ВГУ, Серия : Системный Анализ и Информационные Технологии. – 2010. – №1. – С. 12-14.
- 21 Матюгин, И. Ю. Эффективная память: как быстро запоминать цифры и английские слова / И.Ю. Матюгин. – Москва : Рипол классик, 2004.

22 Гиппенрейтер, Ю. Б. Хрестоматия по общей психологии. Психология памяти / Ю. Б Гиппенрейтер. – Москва : Изд-во МГУ, 1998.